



HAL
open science

Vers une naturalisation de la théorie de la vertu : prolégomènes à une psychologie positive de la vertu

Alberto Masala

► To cite this version:

Alberto Masala. Vers une naturalisation de la théorie de la vertu : prolégomènes à une psychologie positive de la vertu. Philosophie. Université Paris-Sorbonne - Paris IV, 2009. Français. NNT : . tel-00431132

HAL Id: tel-00431132

<https://theses.hal.science/tel-00431132>

Submitted on 10 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS IV – SORBONNE
ECOLE DOCTORALE V CONCEPTS ET LANGAGES

THÈSE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ PARIS IV

Discipline : Philosophie

Présentée et défendue publiquement par

ALBERTO MASALA

Le 10 juillet 2009 à Paris

Titre : Vers une naturalisation de la théorie de la vertu

Prolégomènes à une psychologie positive de la vertu

Directeur de thèse : Daniel Andler

Jury

M. Daniel Andler, professeur à l'Université Paris-Sorbonne (Paris IV)

M. Jérôme Dokic, directeur d'études à l'Ecole des hautes études en sciences sociales

M. Edouard Machery, professeur à l'Université de Pittsburgh

M. Patrick Pharo, directeur de recherches au CNRS

Mme Julia Tanney, professeur à l'Université du Kent

The future is here. It is just not widely distributed yet.

William Gibson

Remerciements

Je tiens à remercier :

Mon directeur de recherche Daniel Andler, pour son soutien intellectuel, motivationnel et institutionnel pendant plusieurs années ;

Walter Mischel, pour m'avoir introduit à sa théorie de la personnalité, qui a grandement influencé le présent travail ;

Hichem Naar, Jérôme Ravat et Joana Baguenier pour avoir lu et commenté des versions antérieures du manuscrit ;

à nouveau Hichem Naar, Jérôme Ravat, mais aussi Francesco Callegaro, Riccardo Fanciullacci, Steeves Demazeux et Olivier Jacquemont pour des suggestions importantes et des discussions très fructueuses ;

mon ancien directeur de recherche Pascal Engel, pour avoir encouragé mes débuts dans la recherche et dans l'étude de la théorie de la vertu ;

le groupe Nash (Hugo Mercier, Nicolas Baumard, Olivier Morin et Nicolas Claidière) pour m'avoir introduit à l'approche naturaliste en sciences humaines et à la théorie de l'évolution ;

enfin, ma famille, pour tout le soutien pendant ces années de travail.

Remerciements	3
Introduction. Contexte et motivation du projet.	10
Vers une nouvelle science de la vertu.....	10
Le contexte : D'une crise des fondements à une nouvelle science de la vertu. Conditions de possibilité du dialogue entre psychologie et philosophie	15
Le modèle globaliste « néo-aristotélicien » en tant que représentant de la psychologie de la vertu classique.....	24
Le modèle néo-aristotélicien.....	25
L'ambition morale du modèle « néo-aristotélicien » : la psychologie de l'excellence morale vs. la psychologie de la « décence » morale	30
Les modèles psychologiquement ambitieux et les approches révisionnistes	33
Introduction à la première partie.....	38
La nécessité d'une psychologie des vertus basée sur la psychologie empirique.....	42
L'insuffisance du débat contemporain sur l'adéquation empirique de la théorie de la vertu face à la menace situationniste : la rhétorique de « la philosophie qui nie l'évidence factuelle »	43
Le dialogue avec la psychologie empirique est-il possible ? Philosophie de la psychologie wittgensteinienne et données expérimentales	46
Psychologie philosophique normative	54
Psychologie normative et données empiriques	58
Pourquoi et comment étudier le sens commun ?	64
Introduction et résumé - Le sens commun, ce grand inconnu	64
Section I – l'étude du sens commun en philosophie – l'insuffisance des approches traditionnelles	67
Quelle est la position du sens commun ? L'importance de la philosophie expérimentale	67
« Populisme normatif et expertise » vs « fiabilisme et révisionnisme éclairé ».....	69
La nécessité de la philosophie expérimentale, dans l'approche populiste.....	73
Les philosophes en tant qu'experts dans l'étude des intuitions réfléchies	74
Les philosophes en tant qu'experts des intuitions « robustes »	76
La nécessité de la philosophie expérimentale dans le cadre du fiabilisme.....	77
La psychologie morale est l'étude des engagements dynamiques du sens commun : des intuitions à une jurisprudence naïve	82
Les intuitions sont un produit cognitif complexe	83
L'étude des engagements dynamiques du sens commun	84
Section II - Qu'est une théorie naïve en psychologie ?	87
Extraction d'engagements ontologiques d'aspects locaux du fonctionnement de la cognition	92
Confusion entre le fonctionnement d'un domaine cognitif et ses engagements ontologiques	96
La théorie du caractère implicite du sens commun	100
Introduction et résumé	100
Section I - Cartographie de l'usage conceptuel : Existence et portée des concepts de trait dans le sens commun	104
Attribution	105
Prédiction.....	108
Explication.....	111
Traits et stéréotypes	114

Scepticisme quant au rôle des concepts de trait dans la psychologie populaire.....	116
La réduction entre des concepts universels et importants du sens commun est très improbable.....	118
Dissociation entre traits et croyances/désirs dans l'ontogenèse.....	118
Dissociation entre traits et croyances/désirs.....	119
Statut.....	120
Anti-réalisme global.....	121
Anti-réalisme spécifique aux traits.....	123
Théorie spontanée des traits.....	127
La possibilité d'une science de la personnalité.....	133
Introduction et résumé.....	133
I. Personnalité et évolution : au-delà du réductionnisme.....	135
Expliquer la variation individuelle : la sélection fluctuante.....	135
Personnalité animale.....	141
L'évolution : un cadre méta-théorique pour la structure de la personnalité.....	143
Réductionnisme dans l'histoire des théories de la personnalité.....	148
Le modèle de l'excellence.....	149
Le modèle des valeurs.....	151
II La définition du domaine d'une science de la personnalité.....	153
Délimitation du domaine de la psychologie de la personnalité : la différence entre la psychologie de la personnalité et l'étude de l'identité personnelle.....	154
Introduction deuxième partie.....	159
La psychologie traditionnelle des traits et son inadéquation pour évaluer la théorie de la vertu.....	163
Introduction et résumé.....	163
La nécessité de désambiguïser la psychologie morale de la théorie de la vertu en tant que théorie des traits.....	165
La notion de trait en psychologie.....	168
Relation entre psychologie et théorie de la vertu.....	175
Le problème de l'homogénéité locale et l'interprétation réaliste des traits de caractère dans la tradition philosophique et psychologique.....	176
Les traits généraux sont-ils des artefacts ? Défense d'un point de vue réaliste.....	183
Les effets holistes sont encore moins fiables que ceux associés à un seul trait.....	189
Approches psychologiques de l'étude de la vertu.....	190
La classification VIA.....	194
Approches typologiques et développementales à l'étude de la personnalité.....	199
Introduction et résumé.....	199
Approche typologique à la personnalité et théorie de la vertu.....	201
La nécessité d'une typologie « molaire ».....	206
Typologies « macro » : les adaptés, le sous-contrôlés et les sur-contrôlés.....	209
Approche développementale à la personnalité.....	211
Conceptualiser la stabilité et la continuité de la personnalité dans le temps.....	214
Macro-typologies développementales.....	218
Modèles socio-cognitifs de la personnalité.....	220
Aspects communs des approches socio-cognitives.....	222
Questions ouvertes : la nature des unités cognitives et le rôle des modules dans la cognition sociale.....	226
KAPA (knowledge appraisal personality architecture).....	231
Modèle de l'auto-régulation de Morf.....	232
CAPS (cognitive-affective personality system).....	234

Les formes de cohérence « substantielle » proposées par l'approche socio-cognitive : profil de personnalité, types, dynamiques relationnelles et la personnalité d'une culture	237
La sensibilité au rejet	239
RS et auto-régulation	244
RS et personnalité borderline	245
RS et dépression	246
RS liée au statut social	248
Narcissisme	249
Personnalité dépendante	250
La cohérence de la personnalité à un niveau supra-individuel : les dynamiques dans une relation et la dimension culturelle	251
Introduction troisième partie	256
Premier chapitre	257
Deuxième chapitre – la critique du situationnisme radical	258
Troisième chapitre – Vertu morale et psychologie de l'expertise	260
Chapitre 4 – l'intégration de la vertu et le rôle de la sagesse	262
A la recherche de la cohérence dans le comportement : la leçon de la psychologie de la personnalité pour la théorie des vertus	264
I. L'étude de la cohérence inter-situationnelle (<i>cross-situational consistency</i>) et sa pertinence pour la théorie de la vertu : prémisses méthodologiques	264
Classification des formes différentes de cohérence	264
II. A la recherche de la cohérence inter-situationnelle : 80 ans de psychologie de la personnalité et leur pertinence pour la théorie de la vertu	271
L'enquête sur l'éducation du caractère et d'autres études classiques	271
Funder : une tentative manquée de dépasser le « coefficient de personnalité »	275
Vranas : le plafond de 0,30 est réel mais compatible avec un niveau élevé de cohérence	280
A la recherche de la cohérence perdue	281
Existence de traits non intuitifs ou non représentés dans le langage	282
La perception partagée des situations ne correspond pas à nos attentes	283
Découverte des dimensions générales qui expliquent la similitude entre situations	285
Template Matching	286
Nature subjective des traits	289
Méta-Traits	297
Conclusion	308
Le menace radicale du situationnisme	310
Introduction au problème du pouvoir de la situation	310
I. La littérature situationniste : la force de la situation et l'erreur fondamentale d'attribution	312
Le paradigme de Milgram	313
Les effets sur l'humeur	316
Le paradigme du Bon Samaritain	318
L'influence du groupe	319
Influences positives de la situation : altruisme	320
Dissonance cognitive et « channeling factors »	322
L'erreur fondamentale d'attribution	324
II. Genres de « situationnismes » et « situations » différents en philosophie et psychologie	325
Interprétations causales du « pouvoir de la situation »	328
Interprétations non causales : le rôle d'une référence aux situations dans la prévision du comportement	334
III. Le situationnisme contre la théorie de la vertu	338
La fragmentation radicale du caractère	339

Des traits de caractère conçus de façon « behavioriste » n'existent pas	342
Les traits de caractère n'existent pas : l'incompatibilité entre l'existence du caractère et de la vertu et les résultats situationnistes qui concernent le comportement moral.....	345
Les traits de caractères moraux sont extrêmement rares	347
L'étiologie de l'erreur : l'erreur fondamentale d'attribution.....	350
Formation d'impressions stéréotypées	352
Le maintien d'impressions stéréotypées dans l'approfondissement d'une connaissance ...	354
Conclusion	361
La vertu et la psychologie de la compétence et de l'expertise	362
Psychologie de la compétence et de l'expertise	364
Compétence et vertu dans la tradition philosophique.....	370
La psychologie de la compétence d'après Dreyfus	374
Evaluation du modèle intuitionniste : le rôle de la réflexion.....	376
La notion d'automatisme en psychologie.....	377
Contexte de déclenchement	377
Profil d'automatisme	378
Le déclenchement automatique de l'action vertueuse.....	378
La complexité des processus cognitifs automatiques	380
Les limites de l'automatisation : le rôle de la réflexion.....	384
Flexibilité et excellence	387
L'intégration des vertus et le rôle de la sagesse	390
Introduction	390
SECTION 1 : brève introduction à la notion de sagesse	391
La notion de sagesse : questions méthodologiques	391
La double nature de la sagesse : le « jugement sage » et la « personnalité du sage »	393
Le noyau sémantique minimal de la sagesse en tant que jugement sage.....	395
La sagesse en tant que personnalité vertueuse du sage.....	398
Le rôle unificateur de la sagesse dans la psychologie classique des vertus.....	400
Différentes formes de sagesse	404
Études empiriques du concept de sagesse du sens commun.....	408
Cohérence dans l'usage.....	409
Validité	411
Variation dans l'usage.....	412
Section II évaluation du bien fondé empirique du rôle unificateur de la sagesse.....	412
Approches empiriques à la sagesse.....	412
Baltes	413
Sternberg.....	419
Approches empiriques à la sagesse en tant que personnalité.....	424
Conclusion	427
Remarque conclusive : quelle place pour l'excellence du caractère dans la philosophie normative du XXIème siècle ?.....	429
INDEX.....	435
BIBLIOGRAPHIE.....	436

Introduction. Contexte et motivation du projet¹.

Vers une nouvelle science de la vertu

Le caractère est un sujet de réflexion fort courant. Nos proches, nos connaissances, les personnalités en vue, nous les jugeons timides, courageux, loyaux, confiants, généreux, etc. Nous avons aussi des idées sur la façon dont le caractère pourrait être développé, changé quand cela est possible, ou au moins influencé. Un voyage peut nous ouvrir les yeux sur la pauvreté dans le monde et développer notre générosité ; la guerre ou l'activité militaire peuvent renforcer le courage ; l'interaction avec une personne que nous admirons et que nous cherchons à imiter peut augmenter notre confiance. Des réflexions et des croyances de ce genre sont explorées et transmises par le sens commun, dans la culture et dans la sagesse populaires, dans la littérature et les arts.

Toutefois, depuis que l'homme a développé l'ambition d'un savoir plus puissant, justifié et systématique (c'est-à-dire depuis que l'homme a développé de véritables *théories*), le fonctionnement et le développement du caractère ont fait aussi l'objet de ses efforts explicitement théoriques. Notamment, ici nous nous intéressons aux *théories prescriptives* qui proposent un idéal de personnalité vers lequel on devrait converger. Le caractère idéal est qualifié de « vertueux », du latin *virtus*, qui traduit le terme grec pour désigner l'excellence : « *arété* ». Les théories normatives du développement du caractère vertueux (ou, plus simplement, théories de la vertu) ont plusieurs millénaires d'histoire : on peut les faire remonter aux débuts de la philosophie en occident (les théories de la vertu de Platon et Aristote²) et en Chine (la théorie de la vertu de Confucius³).

Cette thèse s'inscrit dans cette tradition théorique et vise à signaler et à encourager un tournant majeur dans son évolution : la naissance d'une véritable science interdisciplinaire de la vertu, qui prendrait la place d'une théorie philosophique de la vertu déconnectée de la recherche en psychologie empirique. Le but de ce travail est de montrer la pertinence pour un renouveau de la théorie de la vertu d'un certain nombre de traditions expérimentales en psychologie de la personnalité et en psychologie sociale. Pour le dire plus simplement, il s'agit de montrer la contribution possible de la psychologie contemporaine à une philosophie normative du caractère

¹ Cette introduction motive le projet de la thèse et le situe dans le contexte de la recherche contemporaine sur le caractère. Pour un véritable résumé de la thèse, voir les introductions aux trois parties.

² Aristote, *Ethique à Nicomaque*, traduit par J. Tricot (J. Vrin, 1972).

³ Eric L. Hutton, "Character, Situationism, and Early Confucian Thought," *Philosophical Studies* 127 (2006).

moral, c'est-à-dire une théorie qui nous prescrit la façon dont nous devrions développer notre caractère.

Toute théorie normative du caractère ne constitue pas une « théorie de la vertu ». Nous proposons la définition suivante :

« la théorie de la vertu est une entreprise théorique cherchant à élaborer une théorie normative du caractère basée sur l'exploitation du potentiel normatif des concepts « épais » de trait⁴ qui existent dans toute langue connue ».

Cette définition est minimale, mais permet d'opérer une distinction entre théorie de la vertu et théorie normative du caractère. En principe, une théorie normative du caractère pourrait être construite à partir de n'importe quelle notion psychologique, même très contre-intuitive. Par exemple, on pourrait soutenir que l'harmonie de la personnalité réside dans l'équilibre entre le fonctionnement des hémisphères gauche et droit du cerveau. Ce serait une théorie normative du

⁴ Les termes de trait « épais » sont une classe spécifique de termes qui existe dans toute langue connue (voir le chapitre consacré à la théorie des traits du sens commun). Ils sont utilisés pour qualifier des personnes ou des actions : « courageux », « honnête », « loyale », « magnanime », « prudent », etc. Il s'agit de termes évaluatifs *sui generis*, associés à un contenu descriptif psychologique très riche. Pour cette raison ils sont dits « épais », par opposition à des termes évaluatifs qui ont un contenu descriptif psychologique plus pauvre, comme « correct » : le fait de savoir qu'une action a été qualifiée de correcte ne nous donne pas davantage d'informations psychologiques sur la nature de cette action. Par contre, le « courage », par exemple, est associé à un certain type d'attitude et à des réactions émotionnelles spécifiques, chez l'individu courageux. Même si les termes de trait épais sont évaluatifs (ex : le courage est une qualité positive), leur contenu descriptif est « unique » : on pourrait douter de la valeur du courage, mais il n'existe pas d'autres concepts neutres pour désigner les mêmes propriétés psychologiques auxquelles on se réfère par le terme « courageux ». Cette propriété explique leur importance en psychologie morale : en ignorant les termes de trait épais on perdrait toute information substantielle sur la psychologie morale cautionnée par le sens commun. Il s'agit littéralement d'une ressource théorique. Pour la distinction entre termes épais (*thick*) et termes « minces » (*thin*), voir B. Williams, *Ethics and the Limits of Philosophy* (Cambridge, Mass: Harvard University Press 1985).

caractère entièrement basée sur des connaissances scientifiques. La théorie de la vertu garde un lien privilégié avec le vocabulaire évaluatif du sens commun.

Pourrions-nous adopter une définition plus substantielle ? Non, si nous voulons garder l'unité d'un débat qui a des millénaires d'histoire. Cette définition est certes minimale, mais elle permet de construire un dialogue et une confrontation critique de Confucius jusqu'aux analyses de la vertu dans la psychologie empirique contemporaine, en passant par les classiques de la philosophie occidentale et par l'éthique de la vertu contemporaine. Ce qui rassemble des auteurs venant d'horizons théoriques aussi disparates est la volonté d'exploiter de la façon la plus complète et satisfaisante possible le potentiel du « vocabulaire épais », dans le but de formuler une théorie normative du caractère.

Au premier abord, le thème de la vertu semble être substantiel et clairement identifiable. Les vertus sont le courage, l'honnêteté, la sagesse, la générosité, l'ouverture d'esprit, etc. Depuis qu'il y a des témoignages écrits, on en parle, et les philosophes actifs dans n'importe quelle tradition théorique les ont toujours mises en valeur et analysées. Aujourd'hui comme dans l'antiquité, une théorie de la vertu devrait nous donner une explication de leur fonctionnement, de la manière dont on les acquiert, on les développe et on les perfectionne, de leur utilité pratique et des rôles normatifs qu'elles jouent.

Entreprise ambitieuse mais au moins bien définie. Cette apparence est trompeuse. Une fois dépassé le sentiment superficiel de familiarité toujours associée aux notions proches du sens commun, l'un des problèmes principaux d'un débat sur la vertu de deux millénaires s'avère être la définition même de la question. Le noyau sémantique minimal de la notion de vertu exprime le caractère admirable d'actions, de dynamiques psychologiques ou de personnes. Mais dès qu'on dépasse le niveau superficiel, on se rend compte que la vertu a été associée à une quantité tellement hétérogène de propriétés psychologiques, de rôles normatifs et de modèles moraux (censés l'exemplifier) qu'on peut légitimement se demander si on a toujours discuté du même sujet.

Une explication du courage comme expertise dans la maîtrise de soi face au danger physique et une deuxième explication qui interprète le courage comme un trait nécessaire pour gérer toute interaction sociale avec succès (au travail, dans la vie privée) sont-elles en opposition ? Ne se réfèrent-elles pas plutôt au courage militaire et au courage social ? Peut-être sont-elles compatibles et même intégrables, mais alors une nouvelle question se pose : dans quelles circonstances deux théoriciens de la vertu s'opposent-ils ? Dans quelles circonstances discutent-ils de problèmes différents ? La richesse d'options et la polysémie de la notion de vertu soulèvent la menace d'une

fragmentation du débat. Pourtant, nous croyons qu'il vaut la peine de chercher ce qui pourrait faire l'unité.

Le caractère problématique d'un recrutement pour des fins théoriques des notions du sens commun se manifeste pleinement dans le débat sur la vertu. Les théoriciens de la vertu n'ont pas su trouver un accord sur les questions les plus élémentaires : les termes de trait expriment-ils des propriétés de traits de caractère (ou en tout cas de structures psychologiques bien précises) ou d'actions ? Le courage ou la générosité sont-ils avant tout des propriétés d'actions ? Des théoriciens tel que Hurka l'ont pensé⁵. Dans ce cas, d'un point de vue de la structure du caractère, la notion de courage serait non informative : le courage (en tant que trait) serait à identifier avec n'importe quelle disposition stable à agir de façon courageuse (voir ci-dessous). La tradition des philosophes qui n'associent pas leur théorie de la vertu à une psychologie morale substantielle est beaucoup plus développée qu'on ne le pourrait croire.

Une majorité de théoriciens s'accorde cependant sur le fait que les termes de trait se réfèrent sans doute à des structures psychologiques de l'agent, et non avant tout à des actions. Le courage serait un trait de caractère. Il est question ici d'une psychologie morale informative : mais il n'y a pas d'accord sur le fonctionnement psychologique des vertus, et, en cas d'accord, il y a désaccord sur les rôles normatifs associés aux mêmes structures psychologiques.

Il y a deux variables fondamentales dans toute théorie de la vertu : la psychologie morale des vertus⁶ et les rôles normatifs qui sont associés aux structures psychologiques définissant la vertu. Dans le cadre de la théorie de la vertu, cette distinction est associée à deux débats qu'il est important de distinguer. La distinction semble artificielle : les vertus ne sont-elles pas *définies* par leurs propriétés normatives ? Le courage, la justice et la sagesse ne sont-ils pas précisément les traits de caractère qui garantissent l'épanouissement et le bonheur, des traits définis par le fait qu'on devrait les développer ? Bien sûr, il n'y a pas de pure psychologie des vertus qui ne serve pas aussi de base à la formulation de prescriptions normatives, mais des philosophes qui acceptent la même psychologie des vertus sont en désaccord sur la nature exacte des rôles normatifs joués par le courage, la justice, etc. D'après certains philosophes, la notion de vertu permet de définir la notion d'action moralement correcte : « l'action correcte correspond à ce que le vertueux ferait dans la même situation »⁷. D'autres considèrent que la vertu se limite à nous aider à agir de façon correcte,

⁵ T. Hurka, "Virtuous act, virtuous dispositions," *Analysis* 66, no. 289 (2006). Voir prochain chapitre pour un approfondissement de cette question.

⁶ Quelle est la nature psychologique de la vertu? Quels sont les états et les processus psychologiques qui sous-tendent une conduite vertueuse? Comment les vertus interagissent entre elles? Etc.

⁷ R. Hursthouse, *On Virtue Ethics* (Oxford University Press, 2000).

la notion d'action correcte étant définie par d'autres moyens. À cause de ce désaccord quant aux rôles normatifs de la vertu, il est certainement possible d'isoler la question du fonctionnement psychologique des vertus (psychologie morale des vertus) du problème des rôles normatifs qui leur sont associés.

Le présent travail se concentrera sur le premier débat psychologique aux dépens du deuxième débat normatif. Nous allons analyser la psychologie morale des vertus, sans avoir l'espace (malheureusement) pour trancher sur les conséquences normatives du modèle proposé. Il est évident que le modèle que nous allons retenir (spécialisation dans un nombre limité de vertu) mettra à mal les ambitions normatives qui ont été plus naturellement associées à d'autres modèles psychologiques des vertus (unité des vertus : maîtrise de toutes les vertus en même temps). Toutefois, l'exploration rigoureuse de la marge de manœuvre normative qui reste ouverte à partir de l'acceptation de notre modèle psychologique demanderait une autre thèse. Nous nous limiterons à esquisser quelques scénarios normatifs dans la conclusion. En ce sens, il s'agit d'une thèse sur la psychologie morale (des vertus), non d'« éthique de la vertu ».

Nous avons déjà introduit la notion de théorie de la vertu. Qu'est-ce que alors l'éthique de la vertu ? Il s'agit d'une version très ambitieuse de théorie de la vertu, qui a occupé le devant de la scène une première fois dans l'antiquité et une deuxième fois dans le débat des dernières décennies. Il ne fait guère de doute qu'Aristote a défendu une éthique de la vertu, et il est ainsi pour la plupart des théoriciens de la vertu actuels⁸.

L'éthique de la vertu est ambitieuse dans le cadre du deuxième débat qu'on vient d'évoquer : le débat sur les rôles normatifs associés aux vertus. Imaginons qu'il y ait un consensus sur une psychologie morale des vertus substantielle (premier débat) : quel rôle vont jouer ces vertus dans une théorie normative ? Ce sera uniquement question de proposer un modèle d'individu vertueux à imiter ? Allons-nous prescrire de façon systématique l'apprentissage de ces vertus ? Quelle est la fonction des vertus ? Les vertus nous aident-elles à devenir plus moraux ? Allons-nous soutenir que les vertus *définissent* la dimension morale et ses concepts de base (la notion de devoir, d'action moralement correcte, de bonheur, etc.)

⁸ A l'origine du regain d'intérêt pour la notion de vertu après la deuxième guerre mondiale il y a un renouveau de l'éthique de la vertu, notamment d'inspiration aristotélicienne. Les noms les plus célèbres sont: Elizabeth Anscombe, Philippa Foot, Peter Geach, Rosalind Hursthouse, Alasdair MacIntyre, Bernard Williams, John McDowell, Anthony Kenny, Sabina Lovibond et Georg Henrik von Wright. Comme nous l'expliquons dans le chapitre « La nécessité d'une psychologie des vertus basée sur la psychologie empirique », l'ambition normative de l'éthique de la vertu contemporaine n'a pas été associée à l'ambition de formuler une théorie psychologique informative des vertus. L'éthique de la vertu ancienne était caractérisée par l'union de ces deux ambitions.

Même s'il serait difficile de trouver une définition consensuelle⁹, l'éthique de la vertu se place sans doute du côté ambitieux de ce continuum. Elle est souvent identifiée à la thèse selon laquelle les notions de devoir, de bien et de juste doivent être expliquées par la notion de vertu. Si tous les théoriciens ne s'accordent pas sur cette question précise, il y a sans doute une tendance à placer la vertu au centre de la structure d'une théorie normative.

Or, il est tout à fait possible de défendre une *psychologie* morale des vertus complexe et fascinante sans souscrire à l'éthique des vertus. Par exemple, à l'instar de Kant¹⁰, en soutenant que l'étude des vertus correspond à une partie de la morale, et que les vertus nous aident seulement à devenir moraux, sans définir la dimension morale.

Le présent travail ne s'occupe pas du débat sur les rôles normatifs des vertus, qui oppose la théorie de la vertu (modeste) à l'éthique de la vertu (ambitieuse). Le débat a fait couler beaucoup trop d'encre, car, pour l'instant, nous ne disposons même pas d'une psychologie morale des vertus consensuelle. C'est un problème plus urgent qui doit être résolu en premier. Ensuite, comme nous l'avons dit, une fois que les bonnes contraintes psychologiques seront en place, l'exploration de la marge de manœuvre normative demandera tout autant de travail.

Même dans le cadre de ce panorama théorique fragmenté et étendu dans le temps et dans l'espace, les théoriciens de la vertu convergent vers le but de rendre compte et d'exploiter au mieux le potentiel du vocabulaire normatif du sens commun. Le théoricien de la vertu se comporte face au vocabulaire des traits (« courageux », « honnête », « loyal », « généreux », etc.) comme un état face à une ressource stratégique (diamants, pétrole, mines) qui sera vraisemblablement cruciale pour son développement : il s'agit seulement de trouver la façon de l'exploiter au mieux, mais il ne fait pas de doute qu'il faudra le faire d'une façon ou de l'autre. Le vocabulaire des traits est la ressource stratégique, en vue du développement d'une théorie normative du caractère. Le lien entre les deux, c'est là le travail du théoricien de la vertu.

Le contexte : D'une crise des fondements à une nouvelle science de la vertu. Conditions de possibilité du dialogue entre psychologie et philosophie

L'analyse que nous allons entreprendre est beaucoup plus spécifique et limitée que l'ensemble de pistes et d'idées évoquées par le thème d'un dialogue entre philosophie et psychologie au sujet du

⁹ M. C. Nussbaum, "Virtue Ethics: A Misleading Category?," *The Journal of Ethics* 3, no. 3 (1999).

¹⁰ N. Sherman, *Making a necessity of virtue* (Cambridge University Press New York, 1997).

caractère. Nous n'allons pas dresser un état des lieux de tous les liens envisageables à l'heure actuelle entre pensée psychologique (au sens le plus large) et pensée philosophique (au sens le plus large) sur le caractère. Notre projet sera soumis aux limitations suivantes :

Du côté de la psychologie, nous allons nous intéresser à un nombre limité de traditions contemporaines (plus ou moins récentes) qui ont entrepris une étude *systématique et fortement expérimentale* de la personnalité, et ce choix exclut une masse indéfinie d'idées, d'approches et d'intuitions moins systématiques et/ou qui n'ont pas donné lieu à une activité expérimentale soutenue.

Du côté de la philosophie, nous allons évaluer un modèle normatif du caractère d'inspiration aristotélicienne : ce modèle, qui se définit comme « globaliste »¹¹, propose le développement de traits de caractère vertueux définis par une très grande cohérence inter-situationnelle (le trait se manifeste de façon cohérente dans *toutes* les situations pertinentes : courage à la guerre, dans la vie politique, en famille, etc.) et un lien de renforcement mutuel (être courageux permet d'être encore plus honnête, être loyal permet d'être encore plus généreux, etc. En ce sens, les vertus se renforcent mutuellement).

Même si ce modèle, inspiré d'Aristote, est encore aujourd'hui largement majoritaire, il laisse la place à un grand nombre de conceptions différentes de la vertu.

Nous envisageons un dialogue entre d'un côté certaines traditions expérimentales spécifiques en psychologie de la personnalité, psychologie sociale et psychologie de l'expertise, et de l'autre un modèle de la vertu déterminé, d'inspiration aristotélicienne. Le choix paraît très spécifique et semble relever de nos goûts théoriques personnels.

Mais l'angle d'analyse restreint ne dépend pas uniquement de la limitation de nos forces, il répond en même temps à une nécessité stratégique et historique. Il doit être justifié et replacé dans le contexte de la réflexion théorique contemporaine sur le caractère. Tel est l'objectif de la présente introduction.

À notre connaissance, c'est la première fois qu'une telle quantité de données empiriques est portée à l'attention des philosophes. La seule exception est constituée par le travail de John Doris¹², qui néanmoins nous a offert une sélection plus restreinte et surtout moins représentative du point de vue

¹¹ John M. Doris, *Lack of character: personality and moral behavior* (Cambridge, U.K. ; New York: Cambridge University Press, 2002) 22. Voir prochain chapitre, consacré à la présentation de ce modèle.

¹² Ibid.

des psychologues. Etant donné la popularité du thème du caractère chez les philosophes et le poids académique et institutionnel de la psychologie de la personnalité, ce manque de dialogue manifeste un blocage qui ne peut être expliqué uniquement par des raisons de sociologie de la connaissance (isolement des départements de philosophie et psychologie). En réalité, le manque de dialogue s'explique par une phase de crise des fondements qui a caractérisé la pensée sur la personnalité au XX^{ème} siècle, toutes disciplines confondues. À cause de cette crise, la philosophie et psychologie n'auraient pas pu dialoguer véritablement sur des bases solides.

Après une longue préhistoire présystématique et pré-scientifique qui a duré des millénaires, la pensée théorique sur le caractère humain vit en ce moment un tournant qui annonce la naissance d'une « science de la vertu » ou « psychologie positive de la vertu ». Cette sortie de crise est rendue possible par trois facteurs : (1) une meilleure et plus puissante modélisation de la personnalité en psychologie, libérée des contraintes imposées par l'emprunt du concept de trait et de type au sens commun ; (2) la naissance et la reconnaissance de la « psychologie positive » et (3) une interdisciplinarité forte entre psychologie et philosophie dans le but d'étudier la vertu.

(1) Affranchissement de l'influence du sens commun : en ce qui concerne la théorisation et modélisation du caractère, l'expression « crise des fondements » doit être comprise au sens précis qui lui est associé en philosophie des sciences. Il s'agit de l'impossibilité de faire avancer la pratique scientifique si l'on garde des concepts de base qui s'avèrent être trop flous et ambigus pour une systématisation cohérente. Les concepts fondamentaux de la théorie de la personnalité (trait, type), aussi bien en philosophie qu'en psychologie ont été empruntés au sens commun. Au fil d'un long débat qui a été explicitement vécu comme une crise des fondements, la psychologie de la personnalité s'est progressivement libérée de cet héritage lourd, aboutissant à une modélisation plus technique et précise des différents niveaux de variation individuelle constituant dans leur ensemble la personnalité humaine.

(2) Naissance de la psychologie positive : une modélisation plus puissante de la personnalité ne peut être à la base d'une étude scientifique de la vertu que si la notion d'étude scientifique du fonctionnement *optimal* (par opposition aussi bien au fonctionnement pathologique que normal) de la psychologie humaine est acceptée. C'est là la contribution du mouvement récent de « psychologie positive »¹³. Traditionnellement, la psychologie s'est toujours avant tout occupée soit du fonctionnement pathologique soit du fonctionnement normal de la psyché humaine, délaissant,

¹³ C. Peterson, *A primer in positive psychology* (New York: Oxford University Press, 2006).

entre autre, l'étude systématique (cause, contextes de déclenchement, conséquences, corrélats) d'émotions positives telles que la joie, la gratitude, l'admiration, l'intérêt¹⁴. Aujourd'hui, si le manque d'attention pour tout facteur spécifique qui contribuerait au bien-être humain n'est plus excusable¹⁵, la psychologie hésite encore à proposer des théories globales du fonctionnement cognitif optimal de l'homme ou (encore pire) du « bonheur ». Proposer un modèle global de fonctionnement émotionnel et cognitif excellent serait privilégier une certaine conception de la vie bonne : à l'aune du pluralisme de la société contemporaine, cela est inacceptable. Pourquoi faudrait-il être un saint ? Après tout, n'a-t-on pas le droit d'être un poète maudit ? En tout cas, la psychologie, depuis longtemps déjà accusée d'avoir imposé un modèle de normalité, n'aurait pas vocation à proposer aussi un idéal d'excellence.

Petit à petit, la psychologie est en train d'interpréter de façon moins stricte son statut neutre d'un point de vue normatif. Le souci du respect de la neutralité axiologique n'est pas une raison suffisante pour arrêter les efforts des psychologues positifs. Si plusieurs conceptions de la vie bonne sont possibles, toutes les propositions concevables n'ont pas le même degré de plausibilité. Prenons un exemple. Il y a ceux qui aiment aller en vacances à la mer et ceux qui préfèrent la montagne, mais très peu de gens voudraient passer leurs vacances dans un lit d'hôpital ou en prison. On peut dresser une liste de destinations plausibles tout en respectant une certaine neutralité.

De la même façon, dans le cadre d'une science de la vertu, énumérer une série de scénarios possibles pour le développement du caractère humain et se concentrer sur ceux qui paraissent plus plausibles n'est pas se rendre coupable d'une prise de position contestable. Le poète maudit pourra renoncer à appliquer les techniques qui sont proposées aujourd'hui par les psychologues positifs pour augmenter la qualité et l'intensité des émotions positives¹⁶. L'étude de ces techniques reste bien motivée, car augmenter la positivité émotionnelle est une stratégie plus plausible que d'autres pour se rapprocher de l'excellence du fonctionnement humain ; plusieurs conceptions de la vie bonne pourraient s'approprier ces découvertes, qui, en ce sens, gardent un certain degré de neutralité. En d'autres termes, la psychologie positive peut formuler des prescriptions conditionnelle de la forme suivante : « Si une disposition très forte à éprouver fréquemment des émotions positives vous intéresse, alors vous pouvez appliquer ces techniques ». En formulant de telles prescriptions, les psychologues ont le droit de s'intéresser aux cas où « l'antécédent » se

¹⁴ L. Fredrickson Barbara, *Positivity* (New York: Random House, 2009).

¹⁵ D'un point de vue méthodologique, si on est en mesure d'étudier le contexte de déclenchement et les corrélats de la peur ou de l'anxiété, on sera aussi en mesure d'étudier l'admiration ou la gratitude. Le fonctionnement des émotions positives est un objet d'étude aussi neutre que celui des émotions négatives.

¹⁶ Fredrickson Barbara, *Positivity*.

réfère à un aspect plausible du bien-être humain. Une théorie psychologique de la vertu rentre dans ce cadre méthodologique.

(3) Interdisciplinarité forte entre philosophie et psychologie : même si les craintes des défenseurs de la neutralité axiologique de la psychologie sont exagérées, le mandat et la marge de manoeuvre de cette discipline empirique dans le terrain du prescriptif et normatif sont effectivement limités. De la même façon qu'une force d'interposition dans une zone de conflit a une liberté d'intervention limitée et que son action est strictement contrôlée, la psychologie, en s'aventurant dans le terrain du prescriptif, doit mesurer chaque pas. La psychologie positive, si on la conçoit comme simple branche de la psychologie empirique (à ceci près qu'elle étudie le fonctionnement optimal de l'esprit humain), n'a ni les compétences ni le mandat pour prendre le relais des ambitions traditionnelles de la philosophie en ce qui concerne le caractère. Elle peut avancer un modèle de personnalité « heureuse » ou « épanouie », mais seulement si ces termes gardent une signification vague et proche du sens commun. Le critère du bonheur ou du fonctionnement optimal doit être plausible, non « risqué » et acceptable par tout le monde.

Si cette stratégie consistant à éviter tout ce qui pourrait causer des problèmes paraîtra salubre à certains, elle anéantit a priori toutes les ambitions théoriques de la théorie de la vertu. Au contraire, il faudrait tester ces ambitions et leur donner une première vraie chance, et ne les abandonner que pour des raisons importantes et précises. Les psychologues de la vertu doivent se tourner vers la philosophie beaucoup plus qu'ils ne l'ont déjà fait : la science de la vertu naissante doit être bâtie sur la base d'une interdisciplinarité complète entre psychologie et philosophie. Cette nouvelle science de la vertu aura les compétences (empiriques et normatives) et le mandat pour proposer un modèle psychologique de l'excellence humaine qui soit aussi ambitieux que possible (mais pas plus ambitieux que les données ne le justifient).

Pour cette raison, il est important de viser le plus haut possible au départ. Tester l'adéquation empirique du modèle globaliste d'inspiration aristotélicienne est la bonne façon de procéder. Ce modèle de la vertu n'est pas uniquement le plus répandu et commenté, il est aussi le plus ambitieux en deux sens. Il est le plus ambitieux dans son contenu et d'un point de vue psychologique. D'un côté, il avance l'idéal d'un caractère parfaitement harmonieux qui produit des actions appropriées dans chaque circonstance requise, et dans lequel tous les traits vertueux se soutiennent mutuellement. De l'autre côté, en étant inspiré par les théories du développement de l'excellence et de l'expertise qui étaient répandues à l'époque, le modèle aristotélicien intègre une théorie particulièrement riche de l'habitude et de la formation des réponses émotionnelles adéquates à chaque situation. Autrement dit, ce modèle nous promet des résultats exceptionnels par des moyens

psychologiques informatifs qu'il prend soin de détailler¹⁷. La vertu serait une forme d'excellence pratique, et une théorie substantielle de l'excellence est proposée.

C'est une approche qui s'expose beaucoup et possède un contenu empirique très riche. Il s'agit de deux qualités remarquables. Comme on le verra dans la suite, les théories de la vertu contemporaines qui ont abandonné le modèle aristotélicien ont souvent le défaut d'être peu ambitieuses et/ou peu informatives d'un point de vue psychologique. La difficulté à donner un fondement psychologique précis au fonctionnement de la vertu explique la tendance à formuler des définitions fonctionnelles des vertus, où on nous apprend « ce que la vertu est censée faire mais non comment elle est censée le faire »¹⁸.

Dans d'autres cas, développer un caractère totalement harmonieux paraît un objectif tellement difficile à réaliser que les philosophes y renoncent a priori, en proposant une conception de la vertu beaucoup plus modeste : la vertu serait le fait d'un simple processus de socialisation réussi. La théorie de la vertu décrirait la psychologie du *gentleman*, non de l'homme d'exception (voir prochain chapitre sur cette question). Une modestie injustifiée n'est pas plus acceptable qu'une ambition injustifiée. Le niveau d'ambition justifiée ne peut émerger que d'une évaluation des données et des arguments.

Le présent travail se propose deux objectifs : évaluer le modèle de vertu globaliste aristotélicien à la lumière de nos meilleures connaissances en psychologie de la personnalité et en psychologie sociale, et – condition préalable - établir les conditions de possibilité d'un dialogue entre psychologie et théorie de la vertu. La préparation de l'examen des ambitions classiques des philosophes sera complexe et occupera deux parties sur trois de la thèse. Pour établir les bases de l'interdisciplinarité nécessaire à la nouvelle science de la vertu, il faudra trouver un terrain commun entre deux traditions qui sortent en même temps d'une crise des fondements. Les blocages à dépasser sont nombreux et enchevêtrés : ils tiennent à des facteurs aussi hétérogènes que de « mauvaises » philosophies de la psychologie, les emprunts complexes et mal compris de toute théorie du caractère aux notions vagues du sens commun (traits, types), une compréhension

¹⁷ Voir le prochain chapitre, qui revient sur cette question. La théorie d'Aristote est par ailleurs ambitieuse en un troisième sens: elle associe un grand nombre de rôles normatifs aux vertus. Il s'agit d'une éthique de la vertu (voir ci-dessus). Nous ne nous intéressons pas à ce débat et à ce troisième sens d'ambition: en ce qui nous concerne, le rôle des vertus pourrait bien être uniquement de nous aider à vivre de façon plus morale. Même dans ce cas de figure, la psychologie morale d'inspiration aristotélicienne resterait ambitieuse dans les deux sens indiqués.

¹⁸ Liam Murphy Monahan, "Meeting Anscombe's Demand: Toward a Moral Psychology of Character" (University of Notre Dame, 2004).

inadéquate de la position du sens commun en elle-même, les difficultés « techniques » à modéliser de façon satisfaisante le fonctionnement de la personnalité (complexité du traitement statistique, coût des expériences), etc.

Du côté de la philosophie, il faut sortir de l'influence d'approches métathéoriques qui décrètent a priori l'impossibilité d'un dialogue fructueux avec la psychologie : notamment, on fait référence à des philosophies de la psychologie anti-réalistes d'inspiration wittgensteinienne, et à des conceptions de la normativité qui, même si elles acceptent une interprétation réaliste et causale des processus psychologiques, excluent a priori tout rôle majeur de la psychologie empirique dans le choix des modèles vers lesquels le développement du caractère devrait s'orienter¹⁹.

Même si on admet l'importance de la psychologie, encore faut-il établir les liens possibles entre une masse magmatique de conceptualisations vagues de la personnalité des deux côtés. L'origine du vague tient aux emprunts massifs au sens commun : pendant des millénaires de théorisation ont n'a fait que raffiner et modifier localement les notions vagues et contextuelles de trait et de type. Même lorsque l'apport théorique original était massif, le cordon ombilical n'a jamais été coupé à un endroit clair et de façon nette.

Une modélisation causale rigoureuse n'est pas compatible avec la nature contextuelle des concepts ordinaires. Pour résoudre le problème posé par l'héritage mal assimilé du sens commun, il faut disposer (a) d'une compréhension adéquate de la théorie naïve du caractère et (b) des moyens théoriques et techniques nécessaires pour modéliser de façon satisfaisante la personnalité. Tout cela n'est disponible que depuis quelques décennies (dans certains cas depuis quelques années)²⁰.

Dans une très large mesure, la longue phase préscientifique et la crise des fondements étaient inévitables. Les théoriciens du caractère, toutes traditions confondues, n'auraient pas eu les moyens pour sortir de cette situation jusqu'à il y a un ou deux décennies. La sortie de crise, déjà clairement annoncée et célébrée dans la littérature psychologique, en réalité s'annonce à peine. Malgré la masse considérable de connaissances accumulées par la psychologie de la personnalité depuis un siècle, une véritable science de la personnalité totalement affranchie n'en est qu'à ses balbutiements et, dans une certaine mesure, se cherche encore.

¹⁹ Voir le chapitre « La nécessité d'une psychologie des vertus basée sur la psychologie empirique ».

²⁰ En ce qui concerne l'étude du sens commun, il faut remercier la psychologie de l'attribution et la psychologie interculturelle ; en ce qui concerne les moyens techniques et statistiques, la puissance des ordinateurs, les avancés de la statistiques et les sciences cognitives.

Pour cette raison, la contribution principale de cette thèse consiste en un travail de préparation, de clarification et de déblayage conceptuel nécessaire pour établir le dialogue entre philosophie et psychologie au sujet du caractère.

Dans la troisième partie nous procédons à une évaluation substantielle du modèle globaliste, en arrivant à des conclusions qui nous paraissent solides, notamment le rejet de ce modèle en faveur d'un idéal de vertu en tant que spécialisation morale. Mais les conclusions de cette troisième partie sont modestes, il ne s'agit que d'une ébauche et d'un exemple de pratique de cette science de la vertu que nous préconisons. Même si il est étayé par des arguments solides, l'idéal de vertu en tant que « spécialisation morale » est présenté ici beaucoup plus comme premier résultat d'une science interdisciplinaire de la vertu que comme thèse dont on aurait les ressources (dans ce contexte) de montrer la pertinence dans le débat contemporain en éthique normative. Ce développement étant laissé à une phase ultérieure, le présent travail doit être considéré avant tout comme relevant de la philosophie de la psychologie, et seulement en deuxième instance comme un travail d'éthique. Il s'occupe avant tout de sortir la réflexion normative sur le caractère d'une crise des fondements. La thèse constitue une sorte de « prolégomènes à une psychologie positive de la vertu ».

Notre choix pourrait frustrer quelques sensibilités théoriques, pressées de savoir quelles seraient les véritables conséquences philosophiques de notre modèle (la vertu en tant que spécialisation morale). À cette anxiété, il faut répondre que les questions fondamentales sont celles qui nous hantent tant que nous ne les aurons pas résolues. La hâte de résultats ne peut qu'être contre-productive. Des millénaires de réflexion théorique sur la personnalité ont produit un excès d'hypothèses et de scénarios théoriques sur la nature, le fonctionnement et les relations réciproques des traits de caractère et des vertus. Le problème n'est pas un manque de propositions innovantes, mais un critère pour sélectionner les approches qui sont susceptibles de faire une différence. Trop d'options qui ne peuvent être pas départagées constituent un obstacle pour la mise en place d'un paradigme consensuel qui fasse vraiment avancer le débat.

En se concentrant sur un travail de préparation, cette thèse est en même temps modeste et ambitieuse. La modestie consiste à délimiter clairement la portée de ses conclusions, l'ambition à contribuer à un déblayage conceptuel qui est un passage obligé pour la mise en place d'une science de la vertu. La nécessité de crédibiliser la sortie de crise pour la psychologie de la personnalité justifie aussi notre choix de privilégier (du côté de la psychologie) les approches les plus systématiques et empiriques à l'étude de la personnalité. Seule la rigueur et l'effort de modélisation alimentés par une tradition expérimentale intense peuvent départager certaines pistes dans le magma

d'hypothèses sur la personnalité qui ont été formulées par les psychologues en un peu plus qu'un siècle.

L'espace et l'énergie requis par cet effort de préparation seront importants. Cela limite la possibilité d'avancer sur d'autres fronts complémentaires. Comme nous l'avons dit, l'étude explicite des options normatives associées à la psychologie morale des vertus défendue ici ne sera pas développée. En plus, on ne trouvera dans cette thèse ni une histoire substantielle de la théorie de la vertu ni un traitement de questions « célèbres » (qui ont beaucoup intéressé les philosophes) de psychologie morale des vertus qui sont, d'un point de vue technique, « plus avancées » ou trop de détail. Par exemple, nous ne défendons pas une position sur le lien entre vertus morales et intellectuelles : nous nous concentrons essentiellement sur les vertus morales²¹. Il faut commencer par le début en clarifiant les concepts de base, là où le dialogue avec les savoirs empiriques est le plus facile. Ce sera déjà un travail suffisamment complexe.

²¹ La question du lien entre vertus morales et vertus intellectuelles sera abordée dans le chapitre consacré à la sagesse, paragraphe « différentes formes de sagesse ».

Le modèle globaliste « néo-aristotélicien » en tant que représentant de la psychologie de la vertu classique

Nous avons donné une définition minimale de « théorie de la vertu », pour n'exclure aucun théoricien du débat. La théorie de la vertu est une théorie normative du caractère qui exploite le potentiel normatif des termes de traits. Malgré cela, il faut constater que, d'un point de vue historique, la discussion sur la psychologie morale de la vertu a été largement influencée par une famille de modèles psychologique très ambitieux, qui ont imposé des thèmes et des thèses substantielles. C'est une remarque importante, car, dans le but de préparer un dialogue entre la théorie de la vertu et la psychologie empirique, il faut choisir un modèle psychologique qui représente la tradition philosophique. Le but du présent chapitre est d'exposer la structure du modèle et de justifier notre choix.

Le modèle sera sans doute une abstraction, par rapport aux détails de la psychologie morale défendue par des philosophes déterminés. Cela pourra paraître gênant, mais c'est en fait nécessaire. Choisir d'analyser les détails de la psychologie morale des (par exemple) Stoïciens nous aurait donné l'impression de travailler avec des théories bien réelles, mais cela aurait été prématuré pour deux raisons. Premièrement, aux yeux des philosophes, la recherche en psychologie empirique est généralement considérée comme étant non pertinente (c'est de moins en moins vrai, heureusement) : si l'on n'opère pas une distinction entre les détails d'une théorie philosophique et certains postulats plus généraux qu'elle mobilise, on met une pression exagérée sur la psychologie, ce qui ne favorise pas le dialogue. C'est comme si on demandait aux psychologues de mettre en avant les recherches qui auraient mis à l'épreuve de manière déterminée une certaine théorie stoïcienne, ou alors de quitter la table de négociation. L'attitude selon laquelle les seuls psychologues pertinents sont ceux qui ont étudié *exactement* la même question qui a occupé les philosophes est l'ennemi principal de tout dialogue fructueux.

Deuxièmement, la distinction entre postulats fondamentaux et détails spécifiques est de toute façon nécessaire pour l'évaluation d'une théorie de psychologie morale. Si les postulats de base paraissent problématiques, évaluer le bien fondé empirique des détails n'a pas de sens. Il y a un ordre logique à suivre.

Nous définissons le modèle psychologique des vertus qui sera analysé dans le présent travail « néo-aristotélicien ». Il ne s'agit pas de la théorie de la vertu d'Aristote dans tous ces détails, mais d'un modèle plus général et représentatif de la pensée ancienne sur la vertu. Néanmoins, en ce qui

concerne la nature est les rapports entre les vertus, ce modèle se rapproche davantage de la formulation aristotélicienne, tout en restant fidèle aux postulats plus généraux de la psychologie ancienne des vertus. La formulation aristotélicienne est celle qui est plus apte à engendrer un dialogue fructueux avec la psychologie contemporaine.

Le modèle néo-aristotélicien

La question fondatrice de l'éthique ancienne est « comment vivre ? ». La réponse de la théorie de la vertu est nette : le meilleur investissement pour une vie qui se rapproche le plus possible de la meilleure vie qui puisse être vécue est le développement d'un caractère vertueux. La vertu – conçue comme condition unitaire de l'individu vertueux – est l'excellence du caractère. Quelles que soient les circonstances, un individu vertueux agit de façon opportune et moralement impeccable. La vertu est une forme de savoir vivre, une véritable expertise dans l'art de la vie. Il ne s'agit pas d'une façon de s'exprimer : la philosophie ancienne a littéralement conçu la vertu comme une forme d'expertise et de savoir *sui generis*, portant sur la manière dont il faudrait vivre.

La psychologie morale des vertus des anciens est inspirée par les théories de l'excellence pratique et technique qui étaient répandues à l'époque (voir le chapitre sur la psychologie de l'expertise). Un bon artisan ne se limite pas à fabriquer des chaussures de qualité plus souvent qu'un citoyen lambda ne pourrait le faire. Avoir une capacité supérieure à la moyenne n'est pas suffisant. Un artisan capable fabrique de bonnes chaussures (presque) tout le temps. En termes techniques, sa performance est stable (il peut fabriquer des chaussures aujourd'hui, il pourra les fabriquer dans deux ans) et cohérente. En particulier, il s'agit d'une cohérence inter-situationnelle : dans toutes les situations différentes dans lesquelles ses capacités devraient se manifester, elles s'y manifesteront. Supposons qu'il sache fabriquer des chaussures, des sacs et des gants. Il doit être en mesure de les fabriquer en hiver, en été, en ayant beaucoup de temps à disposition, en étant très pressé et dans toute autre situation raisonnablement compatible avec la pratique de son activité, que les circonstances soient favorables ou défavorables. C'est un modèle d'excellence.

De la même façon, le vrai courage doit se manifester dans la bataille, dans la vie politique, pendant la chasse, et cela qu'on ait le temps de se préparer et que les circonstances soient favorables ou qu'on doive réagir à une urgence dans une situation extrême. Le courage, comme toute autre vertu, doit être caractérisé par une cohérence inter-situationnelle très élevée. L'individu courageux doit toujours manifester son courage lorsque la situation le demande. En ce sens, la psychologie morale classique est ambitieuse : elle avance comme modèle une véritable condition d'excellence.

Mais comment une telle forme d'excellence serait-elle psychologiquement possible ? La théorie de la vertu classique est ambitieuse dans un deuxième sens, elle est ambitieuse en tant que théorie psychologique, car elle croit pouvoir donner une réponse substantielle et informative à cette dernière question.

Le corpus de connaissances pratiques propres à l'expertise de vie de l'individu vertueux a été conçu comme étant très unitaire (voir le chapitre sur la sagesse, paragraphe sur le rôle unificateur de la sagesse). C'est là une dimension holiste fondamentale de la théorie de l'expertise des anciens. On ne peut pas maîtriser un genre de poésie sans maîtriser tous les principes et les formes de l'art poétique. On ne peut pas être expert dans la conduite d'un seul type de bateau sans avoir une compréhension profonde de l'art de la navigation. La compréhension des principes généraux qui règlent un domaine est nécessaire pour atteindre un niveau de performance important. Il est possible d'apprendre des compétences isolées par simple imitation et par expérience, mais la qualité de la performance sera modeste. Seule la théorie nous permet d'atteindre des résultats importants et de laisser derrière nous l'imprécision et les approximations. Mais la théorie qui unifie un domaine d'action doit être assimilée dans sa globalité : il n'y aurait aucun avantage à comprendre uniquement la moitié des principes de l'art poétique.

En ce qui concerne la dimension morale, cette attitude holiste est accentuée : dans une vie moralement excellente, on opère sans doute des arbitrages entre les valeurs et les besoins les plus importants pour un être humain. Partir à la guerre pour défendre sa patrie ne peut être une décision moralement excellente que si elle se base sur une analyse correcte des priorités entre intérêt personnel, familial et collectif. Il faut que l'individu comprenne les tenants et les aboutissants des principaux domaines de la vie. Si l'individu n'attribue pas la juste valeur à sa propre vie et au rôle qu'il peut jouer dans sa famille, il risque de se sacrifier trop facilement. S'il n'attribue pas suffisamment de valeur à l'intérêt collectif, il va se renfermer dans des choix égoïstes. Il semble que l'expertise de vie soit par définition unitaire et globale.

Cependant, une conception trop monolithique du savoir pratique qui serait à la base de la vertu ne facilite pas le dialogue avec la psychologie contemporaine²². La version aristotélicienne de cette

²² Par exemple, Socrate (à partir du Protagoras) et les stoïciens Zenon et Ariston semblent soutenir que les différentes vertus ne correspondent pas à des corpus de savoir distincts. D'un point de vue psychologique, il y a une seule et unique vertu appliquée à des domaines distincts, tout comme la vue, qui est unique, permet de percevoir des scènes différentes. Le courage est la générosité ne seraient que *la* vertu appliquée à des problèmes différents. Ce serait difficile de mettre la psychologie contemporaine en relation avec une telle conception. Voir chapitre sur la sagesse, paragraphe sur le rôle unificateur de la sagesse.

doctrine (selon laquelle la vertu est une forme d'expertise rendue possible par un corpus de savoir pratique très unitaire) maximise la possibilité d'un échange fructueux.

Chez Aristote, on retrouve une forme de *pluralisme* de la vertu : il est possible de diviser le corpus de connaissances intériorisée par l'individu vertueux en un nombre limité de macro domaines de base. Chaque domaine correspond à une vertu spécifique. La détermination des domaines ne correspond pas à la spécificité sociologique d'une culture donnée mais à des problèmes que tous les être humains doivent affronter. Contrairement à certaines évaluations d'Aristote, selon lesquelles l'on ne pourrait appliquer ça conception de la vertu qu'à son propre contexte culturel, sa stratégie est universelle²³. Par exemple, il n'y a pas des vertus spécifiques aux métiers ou aux rôles sociaux (les vertus du commerçant, les vertus de l'avocat, les vertus sacerdotales), mais la vertu du courage est requise dans tous les contextes où il faut gérer des émotions face au danger, que ce soit dans la bataille ou dans la vie publique. La vertu de la justice est requise dans tous les cas où il s'agit de répartir des mérites, des responsabilités et des ressources. Bien sûr, un commerçant ou un avocat devront maîtriser un corpus de connaissances pratiques spécifiques à leur fonction (et, si on le veut, on peut bien donner un sens dérivé à la notion de vertu d'un avocat), mais la vertu concerne des « fondamentaux » du savoir vivre, des classes de problèmes que tous les être humain finiront par avoir, quelle que soit leur culture d'origine.

Les différents corpus d'expertise qui constituent la vertu ne peuvent être appris et maîtrisés qu'en même temps, car (pour les mêmes raisons qui ont poussé la philosophie de la vertu ancienne à envisager une unité encore plus radicale) ils sont complémentaires. Au début de ce processus, on peut avoir des réactions émotionnelles qui vont dans le bon sens, par exemple, ne pas avoir peur face à un ennemi dans la bataille. Il s'agit de ce qu'Aristote appelle vertus naturelles. Il n'y a pas de d'implication mutuelle entre vertus naturelles. On peut avoir une tendance naturelle à la justice et au courage mais non à la générosité. Les vertus naturelles correspondent à de bonnes réactions émotionnelles. Mais elles ont besoin d'être calibrées : même si l'on n'a pas tendance à fuir face au danger, il reste encore un gros travail de dressage émotionnel à faire.

Il faut que la réaction émotionnelle soit parfaitement adéquate et proportionnelle à la gravité de la situation. C'est la doctrine aristotélicienne du juste milieu. Ce n'est pas suffisant que les émotions aillent dans le bon sens : l'action efficace demande un calibrage fin. Face à une personne en détresse, on risque facilement d'avoir trop ou trop peu de compassion. L'individu vertueux aura exactement le niveau de compassion requis dans le contexte.

²³ M. Nussbaum, "Non-Relative Virtues: An Aristotelian Approach," (1993).

« Ainsi également, se livrer à la colère est une chose à la portée de n'importe qui, et bien facile, de même donner de l'argent et le dépenser ; mais le faire avec la personne qu'il faut, dans la mesure et au moment convenables, pour un motif et d'une façon légitimes, c'est la une œuvre qui n'est plus le fait de tous, ni d'exécution facile, et c'est ce qui explique que le bien soit à la fois une chose rare, digne d'éloge et belle²⁴. »

Pour cette raison, le premier composant de l'éducation à la vertu chez Aristote est un dressage fin des réactions émotionnelles. Cela n'est possible que grâce à l'imitation d'individus qui sont déjà vertueux : dans un contexte donné, le jeune doit chercher à réagir de la même façon que l'adulte déjà vertueux.

Mais la théorie de la vertu d'Aristote ne se réduit pas à un dressage des émotions, aussi fin soit-il. Aristote envisage un rôle pour le raisonnement et la délibération. La vie morale est trop complexe, même pour des instincts bien éduqués. Il faut une dimension ultérieure de calibrage, car, dans des cas problématiques, l'individu bien éduqué aura des intuitions divergentes. Ici Aristote fait preuve de réalisme et de plausibilité psychologique. Au lieu de postuler une espèce de sensibilité morale monolithique, il se rend compte qu'il y aura des divergences d'intuitions. A ce stade avancé de la bataille, faut-il infliger une dernière attaque décisive ? Ce serait une preuve de courage ou une manifestation d'injustice, car l'ennemi a été loyal et il mérite d'être épargné ? Ce n'est qu'une capacité de jugement supérieure qui pourra trancher, et cette capacité chez Aristote est la « sagesse pratique » ou prudence (*phronesis*). La prudence est une capacité unique qui permet de gérer des conflits entre des intuitions morales incompatibles. Chez Aristote, les émotions morales ne sont pas uniquement raffinées (on a vu qu'elle doivent être calibrées d'une façon extrêmement fine), elles sont aussi flexibles, car elle doivent être en mesure de s'adapter et de suivre l'arbitrage et les compromis proposés par la sagesse pratique. En ce sens, chez Aristote l'éducation à la vertu produit des habitudes fines, de « haut niveau » et « intelligentes ».

Le dressage des émotions procède en même temps que l'entraînement de la prudence. Le jeune imite les réactions émotionnelles de modèles de vertus, en même temps qu'il s'entraîne à résoudre les premiers conflits qu'il rencontre. Au début, ses performances vont être mauvaises : il va avoir des mauvaises intuitions et il va gérer mal les conflits qu'il perçoit entre ses intuitions. Progressivement, tout les éléments vont se mettre en place, jusqu'au moment où les réactions émotionnelles sont adéquates et la capacité de jugement est mûre. Au fil du développement de la vertu, il y a co-dépendance entre prudence et vertu morales. La capacité de jugement ne peut être bonne que si les émotions sont bien éduquées ; en même temps, les arbitrages passés de la prudence

²⁴ 1108a 26-29 Aristote, *Ethique à Nicomaque*, traduit par J. Tricot (J. Vrin, 2007) 123.

sont nécessaires pour raffiner les réactions émotionnelles futures. Il n'y a de progrès que grâce à un cercle vertueux dans lequel tous les processus s'améliorent grâce à l'amélioration simultanée des autres éléments du système.

« On voit ainsi clairement, d'après ce que nous venons de dire, qu'il n'est pas possible d'être homme de bien au sens strict, sans prudence, ni prudent sans la vertu morale. Mais en outre on pourrait de cette façon réfuter l'argument dialectique qui tendrait à établir que les vertus existent séparément les unes des autres, sous prétexte que le même homme n'est pas naturellement le plus apte à les pratiquer toutes, de sorte qu'il aura déjà acquis l'une et n'aura pas encore acquis l'autre. Cela est assurément possible pour ce qui concerne les vertus naturelles ; par contre, en ce qui regarde celles auxquelles nous devons le nom d'homme de bien proprement dit, c'est une chose impossible, car en même temps que la prudence, qui est une seule vertu, toutes les autres seront données.²⁵ »

Le processus peut ne pas aboutir : c'est le cas de la faiblesse de la volonté (*acrasie* : incapacité à suivre son propre jugement) ou de la condition de qui ne suit son meilleur jugement qu'avec beaucoup d'effort (*encrasie*). Dans une bataille, l'*acratique* n'arrive pas à aider son ami en difficulté, même s'il reconnaît qu'il faudrait le faire, et cela parce qu'il est paralysé par la peur. Dans la même situation, l'*encratique* finit par aider l'ami en détresse, mais au prix d'un effort considérable de maîtrise de soi. Chez ces individus, les émotions ne sont pas suffisamment flexibles pour s'adapter au jugement de la prudence. L'éducation de la vertu est un processus délicat et complexe.

La version aristotélicienne de la théorie de la vertu en tant que forme d'excellence dans l'art de vivre est en même temps pluraliste et unitaire. Elle pluraliste, car Aristote conçoit les vertus morales comme des corpus différentes d'habitudes et d'attitudes émotionnelles. Au niveau psychologique, le courage et la justice sont distincts et peuvent entrer en conflit. En même temps, la sagesse pratique est unique et elle ne peut fonctionner correctement que si les vertus morales sont développées : en ce sens, les vertus s'impliquent et sont nécessairement possédées en même temps.

²⁵ 1144B 30–1145a Ibid. 335-336. Sans éducation et entraînement rigoureux, les vertus naturelles ne sont pas suffisantes pour développer la forme « pleine » des vertus correspondantes (le courage naturel n'est pas suffisant pour être vraiment courageux).

L'ambition morale du modèle « néo-aristotélicien » : la psychologie de l'excellence morale vs. la psychologie de la « décence » morale

La psychologie des vertus du modèle néo-aristotélicien est une psychologie de l'excellence morale, non d'un niveau de moralité respectable, suffisant ou adéquat. Dans le cadre de la philosophie ancienne, cet aspect n'est pas du tout spécifique à Aristote. La théorie de la vertu décrit le fonctionnement psychologique d'une personne exceptionnelle : le sage. L'idée selon laquelle il existe un nombre limité de personnes qui sont caractérisées par un niveau de moralité supérieur et qui devraient être prises comme modèle de conduite par tous les autres se rapproche d'un universel culturel. C'est en tout cas un présupposé considéré comme évident dans les traditions philosophiques orientales et occidentales.

Les théories de la vertu sont en compétition pour expliquer la psychologie du sage : il n'y a pas de doute sur la façon de poser le problème. Dans l'antiquité, en ce qui concerne l'individu sage, tous faisaient référence à une dimension supérieure d'existence et de conduite morale.

Aujourd'hui, cet élément de la définition de la théorie de la vertu ne fait plus de consensus. Cela est dû à l'influence d'une tradition anglo-saxonne en théorie de la vertu, initiée par Hume, qui a littéralement change de sujet²⁶, concevant le but de la théorie de la vertu comme étant celui d'expliquer la psychologie de la « décence » morale. Par décence morale, nous entendons le niveau de moralité d'un sujet pleinement socialisé (avec succès) dans sa communauté. Il n'est pas question d'héros moraux, mais plutôt du citoyen modèle. Pour utiliser un terme qui est en même temps commun et technique dans ce contexte, c'est le « gentilhomme ». Le gentilhomme, au sens technique, est le modèle de la réussite sociale. Quelqu'un parfaitement à l'aise avec les mœurs, les normes et les attentes de sa société, actif avec succès dans la vie privée, dans son travail et au service de sa communauté.

C'est un idéal beaucoup plus modeste et « proximal ». C'est aussi un idéal beaucoup moins utopique. Même s'il a été toujours possible de citer des exemples concrets de sagesse, le profil du gentilhomme est sans doute beaucoup plus répandu est clairement identifiable. Des gentilshommes, il y en a des milliers. En ce sens, l'étude de la psychologie de la décence morale s'apparente beaucoup plus à une démarche descriptive que normative. C'est l'étude de la psychologie d'une socialisation morale pleinement réussie.

²⁶ Cela ne signifie pas que la théorie de la vertu de Hume n'est pas « en compétition » avec les théories ambitieuses qui s'intéressent à l'excellence. Une théorie modeste de la vertu n'est attractive que si les théories radicales sont problématiques.

Ce n'est pas un hasard que cette tradition a été faite remonter à Hume²⁷. Il est reconnu que le philosophe écossais était avant tout intéressé à développer une éthique descriptive : c'est-à-dire la description du phénomène de la moralité humaine tel qu'il existe dans les sociétés connues. Son intérêt dans une éthique normative n'était que secondaire et dérivé. En ce qui concerne cette dernière, Hume est connu pour ne pas être utopiste. Héritier de la philosophie écossaise du sens commun, Hume ne conçoit d'ambition morale plus grande que celle du gentilhomme : la meilleure destinée morale qui nous est promise est d'être parfaitement socialisés d'après les mœurs de notre société. Autrement dit, devenir un modèle moral de référence d'après les critères éthiques du sens commun.

Nous n'avons pas vocation à ouvrir directement un débat pour défendre l'idéal d'excellence morale contre l'ambition à la décence morale (voir brièvement les remarques conclusives pour quelques suggestions sur l'importance de l'excellence). Ici, il nous intéresse uniquement de remarquer qu'une simple socialisation réussie n'est pas le but visé par deux mille ans d'histoire de théorie de la vertu. Au fil du présent travail, nous ne faisons qu'affirmer le principe suivant : les idéaux ambitieux – aussi radicaux soient-ils – méritent d'être évalués avant d'être (éventuellement) rejetés. Nous pensons que le début d'une évaluation des idéaux de vertu de la tradition ne peut qu'avoir lieu aujourd'hui, grâce aux connaissances psychologiques dont nous disposons.

Malheureusement, nous ne pouvons pas nous en tirer aussi facilement que cela. Car les défenseurs contemporains du courant humien « modeste » cherchent à réinterpréter Aristote, soutenant qu'il se serait occupé aussi du gentilhomme, et non du sage²⁸. Cette opération révisionniste ne cible qu'Aristote, car elle ne serait pas envisageable en ce qui concerne Platon ou encore les Stoïciens. Pour ces derniers auteurs, (1) l'insistance sur les capacités exceptionnelles du sage et (2) l'explicitation du fait que la vertu est le résultat du travail d'une vie entière (travail qui ne termine qu'à 50 ou 60 ans, voire jamais²⁹) sont clairement incompatibles avec l'hypothèse selon laquelle il s'agirait d'une description du gentilhomme.

²⁷ Cette interprétation est due surtout au travail de Merritt, voir Maria Merritt, "Virtue Ethics and Situationist Personality Psychology," *Ethical Theory and Moral Practice* V3, no. 4 (2000).

²⁸ L'une des réinterprétations modérées d'Aristote les plus argumentées et détaillées et celle avancée par Sarah Brodie, S. Brodie, *Ethics with Aristotle* (Oxford University Press, USA, 1991).

²⁹ Confucius semble avoir dit : « À 15 ans j'ai commencé à apprendre, à 30 j'étais confirmé, à 40 ans je n'avais plus de doutes, à 50 ans je connaissais la volonté du Ciel, à 60 ans j'étais prêt à l'écouter, à 70 ans je pouvais suivre mes désirs sans transgresser ce qui est juste. » Cité dans Paul Baltes, "Wisdom as Orchestration of Mind and Virtue" (non publié: 2004), 50.

En ce qui nous concerne, cette remarque neutralise déjà en partie l'opération de réinterprétation de la théorie de la vertu. En fait, le modèle « néo-aristotélicien » dont il sera question au fil du présent travail représente une « abstraction » représentative de la psychologie morale classique, susceptible de représenter cette tradition intellectuelle dans un dialogue avec la psychologie contemporaine. Le modèle n'est pas néo-aristotélicien parce que nous nous intéressons avant tout à Aristote, mais parce qu'un modèle d'inspiration aristotélicienne de la psychologie des vertus est le plus apte à entrer en contact avec les théories psychologiques contemporaines. En ce qui concerne la tradition de la théorie de la vertu dans sa globalité, il n'est pas question d'une révision à la baisse de l'idéal proposé.

Toutefois, nous pensons que le modèle « néo-aristotélicien » correspond à la bonne interprétation d'Aristote (même s'il n'inclut pas tous les détails de la théorie aristotélicienne), et il serait étrange que l'auteur censé représenter les théories de l'excellence morale aux yeux de la psychologie contemporaine ne se soit pas intéressé à la question. Est-il plausible de penser qu'Aristote ne se soit occupé que de la « décence » morale ?

D'abord, il faut expliquer pourquoi l'hypothèse ne paraît pas absurde (car elle serait une hypothèse absurde à propos de la théorie des vertus de Platon et des Stoïciens). Aristote ne formule pas une théorie détaillée du processus éducatif censé développer la vertu, au sens du curriculum précis d'études tout au long de la vie exposé par Platon dans la République, ou de véritables exercices spirituels censés forger le caractère des Stoïciens. Il semble se concentrer surtout sur l'imitation de modèles de vertu et sur la pratique : pour devenir courageux, il faut pratiquer le courage sans cesse, suivant l'exemple d'hommes pleinement courageux. Par ailleurs, Aristote ne semble pas insister sur le fait que la vertu est le travail d'une vie entière : il semble croire qu'un individu pourrait être substantiellement vertueux au début de sa vie active adulte³⁰.

Globalement, ces facteurs font penser qu'Aristote se réfère à un normal processus de socialisation réussi : l'éducation n'a pas la rigueur de l'entraînement du sage ; elle semble peu formalisée, naturelle et basée sur la simple pratique ; elle semble terminer dans les délais nécessaires pour le début d'une vie active adulte.

Or, si Aristote s'occupe d'une socialisation normale, il doit être question du « gentilhomme ». Mais cette interprétation ne tient pas la route. D'abord, si Aristote s'intéressait simplement au gentilhomme, il serait le seul philosophe dans son milieu culturel à ne pas se poser la question de la

³⁰ M. Foucault s'est intéressé à ces différentes conceptions de l'éducation, M. Foucault, *Le souci de soi, Histoire de la sexualité* vol. 3 (Paris: Gallimard, 1984).

vertu en tant qu'excellence, et cela tout en mobilisant le même terme que les autres : « *arété* ». Ensuite, Aristote donne comme exemple de vertu une personnalité excellente : Périclès.

« *C'est pourquoi nous estimons que Périclès et les gens comme lui sont des hommes prudents en ce qu'ils possèdent la capacité d'apercevoir ce qui est bon pour eux-mêmes et ce qui est bon pour l'homme en général, et tels sont aussi, pensons-nous, les personnes qui s'entendent à l'administration d'une maison ou d'une cité.*³¹ »

Plus en général, Aristote dans la politique, et suivant l'exemple de Platon, envisage un lien étroit entre la vertu est l'excellence dans la vie publique au service de la société. Pour Aristote, la vertu est l'idéal associé à une classe dirigeante d'exception dans le contexte démocratique d'Athènes.

Or, si Aristote s'intéresse à des hommes politiques d'excellence et non au citoyen modèle, pourquoi ne met-il pas en avant une théorie de l'éducation plus exigeante ? A notre avis, la solution est la suivante³² : il est vrai qu'Aristote ne fait pas peser sur l'individu la contrainte d'un entraînement particulièrement formalisé qui dure toute la vie, mais il envisage un *milieu d'apprentissage idéal*, où le jeune est en contact avec les meilleurs maîtres et modèles moraux et profite d'un processus d'éducation dans le cadre des institutions démocratiques d'Athènes. Dans un contexte idéal, qui prépare au débat, à la réflexion et prône les vertus, le processus d'apprentissage n'a pas à être artificiel et formel : il peut s'agir simplement de suivre le curriculum d'activités et de responsabilité typiques d'un jeune destinée à la carrière politique. Ce parcours peut être suffisamment formateur. Aristote présuppose des *institutions excellentes*.

Contre l'approche révisionniste des interprètes anglo-saxons, nous pouvons donc réaffirmer que le sujet de deux mille ans de théorie de la vertu (Aristote compris) a été bien l'excellence morale du sage, non la « décence » morale du gentilhomme.

Les modèles psychologiquement ambitieux et les approches révisionnistes

La psychologie classique des vertus, représentée par le modèle néo-aristotélicien, vise l'excellence morale. Dans la suite, nous donnerons cette ambition morale pour acquise. Mais dans le présent travail, nos efforts seront consacrés à analyser une autre ambition qui caractérise ce modèle : l'ambition psychologique. Par cette expression, nous nous référons à l'ambition de donner une explication informative du mode de fonctionnement psychologique des vertus. Il est très facile de décrire l'histoire, le rôle et les effets d'un phénomène sans expliquer son mode de fonctionnement.

³¹ 1140b Aristote, *Ethique à Nicomaque* 306. Sagesse pratique (*phronesis*) est traduit par « prudence ».

³² Nous remercions Francesco Callegaro de nous avoir suggérer cette interprétation dans une discussion informelle.

Ainsi, on pourrait tout connaître de l'histoire de l'automobile, des origines jusqu'à la vitesse maximale atteinte par les voitures de formule 1 aujourd'hui, sans savoir expliquer correctement le fonctionnement du moteur. De la même façon, rien de plus facile que louer la vertu et démultiplier les accomplissements de l'individu vertueux (en guerre, dans la vie politique, en famille) sans donner la moindre explication psychologique de la façon dont la vertu fonctionne.

Lorsqu'un théoricien de la vertu ne nous fournit pas une véritable psychologie des vertus, il y a trois cas de figure : (1) tout simplement, il laisse de côté la question, avec ou sans bonnes raisons (2) il ne considère pas que la vertu est un phénomène psychologiquement unitaire, une espèce « psychologique naturelle » : nous considérons comme vertueuses plusieurs attitudes qui n'ont pas grand-chose en commun d'un point de vue psychologique ; (3) il fournit une psychologie des vertus, mais elle résulte trop vague ou exprimée dans un vocabulaire qui n'est pas acceptable pour la psychologie empirique contemporaine : cette théorie ne qualifie pas de vraie explication psychologique d'après nos critères actuels.

Certains théoriciens de la vertu ont des bonnes raisons (depuis leur point de vue) pour ne pas ouvrir la « boîte noire » des vertus (1) : il s'agit par exemple des défenseurs d'une philosophie de la psychologie « anti-mentaliste » d'inspiration wittgensteinienne. Ces auteurs considèrent que les concepts psychologiques (y compris les termes de trait associés aux vertus) expriment des dispositions à l'action (dans un contexte x , l'individu fait y) et ne se réfèrent pas à des processus mentaux « internes³³ ».

Un autre courant de théoriciens contemporains ne se concentre pas sur une psychologie informative des vertus, parce qu'il croit que la vertu n'est pas un phénomène unitaire d'un point de vue psychologique (2). Qu'est-ce que cela signifie ? Tous les théoriciens de la vertu exploitent le potentiel normatif de termes de traits tels que « courageux », « honnête », « généreux », etc. Mais on n'est pas obligé de croire qu'il y ait une explication psychologique unitaire de toutes les attitudes que nous considérons comme « courageuses » ou « honnêtes ». Par exemple, dans un cadre conséquentialiste, Julia Driver définit la vertu comme n'importe quelle habitude qui a des conséquences généralement positives :

« Ma théorie soutient simplement qu'un trait qui a de bonnes conséquences (good-producing) – acquis de n'importe quelle façon – est une vertu morale. Il pourrait être vrai de façon

³³ Voir le chapitre « La nécessité d'une psychologie des vertus basée sur la psychologie empirique », paragraphe « Le dialogue avec la psychologie empirique est-il possible ? ».

contingente que ces traits sont en général ceux qui exigent du travail dur pour être acquis et des efforts pour être maintenus³⁴. »

« Ainsi, la définition correcte de vertu morale sera celle qui a été développée ci-dessus, qui soutient simplement qu'une vertu morale est un trait de caractère qui produit du bien. Soutenir que la vertu morale est un trait de caractère est reconnaître que la vertu morale est quelque chose que ne peuvent posséder que des êtres qui ont un esprit ou un caractère. Ainsi, toute bonne attitude dispositionnelle n'est pas une vertu morale³⁵. »

Les remarques à la fin de la citation sont censées exclure a priori du domaine de la vertu uniquement des dispositions qui n'ont rien à avoir avec les attitudes d'un sujet capable d'action intentionnelle et volontaire, par exemple les réflexes automatiques. N'importe quelle autre habitude qui relève de l'activité cognitive du sujet peut être vertueuse du moment qu'elle a de bonnes conséquences. Si la vertu est définie uniquement par ses conséquences, elle n'est pas un phénomène psychologique unitaire.

Il y a une autre raison possible pour penser que la vertu n'est pas une espèce psychologique naturelle : croire que les termes de vertus se réfèrent avant tout à des actions, non pas à des traits ou à des habitudes. Le sens commun considère sans doute que des actions peuvent être vertueuses. Mais, normalement, les théoriciens de la vertu pensent qu'une attribution de vertu à un acte implique l'attribution du trait vertueux correspondant à l'agent. « C'était un acte courageux » signifie en réalité « cette personne est courageuse (= possède le trait du courage) et l'a manifesté dans le cadre de cette action ». On n'est pas obligé d'adhérer à cette interprétation du sens commun : on pourrait imaginer que les vertus sont des propriétés des actions, et que le sens commun, en attribuant le courage à une action, est *agnostique* sur les habitudes ou les traits de l'agent qui pourraient expliquer pourquoi il est courageux de façon régulière.

Les auteurs qui, comme Hurka³⁶, croient que les termes de trait qualifient avant tout des actions, vont accepter comme étant vertueuse n'importe quelle disposition psychologique qui est à l'origine d'actes vertueux (la notion d'« acte vertueux » étant plus fondamentale que celle de « trait de caractère vertueux » et un terme comme « courageux » ne désignant pas un trait de caractère). De cette façon, ils ne proposent pas une psychologie morale informative.

³⁴ Julia Driver, *Uneasy virtue*, *Cambridge studies in philosophy* (Cambridge, UK ; New York: Cambridge University Press, 2001) 108.

³⁵ Ibid. 107.

³⁶ Hurka, "Virtuous act, virtuous dispositions."

Les deux formes de révisionnisme modéré sont des phénomènes récents dans la longue histoire de la théorie de la vertu. Elles correspondent respectivement à l'abandon de l'ambition d'étudier l'excellence morale et à l'abandon de l'ambition de formuler une théorie psychologique informative des vertus. Même s'il n'y a pas de corrélation parfaite entre les deux formes de défaitisme³⁷, il y a une racine commune : le retour à la théorie de la vertu du sens commun. Tout en exploitant les termes de vertus, la théorie de la vertu classique s'était développée en opposition aux conceptions de la vertu du sens commun, l'idée étant que, une fois qu'on les a bien comprises, on se rend compte que les notions de courage, générosité ou sagesse se réfèrent à un idéal de moralité extrêmement ambitieux³⁸.

Sans cette opération critique, la théorie de la vertu implicite dans le sens commun semble se référer à un ensemble de qualités appréciables dans la conduite humaine. Il est sans doute mieux d'être courageux, honnête et juste, mais, si on s'arrête au niveau de complexité que ces notions ont chez le sens commun, les concepts de vertu ne sont suffisants ni pour définir les propriétés d'un modèle d'excellence morale, ni pour expliciter un modèle psychologique substantiel. Dans le sens commun, les vertus ne semblent être guère plus qu'une liste de qualités que nous apprécions. Ce n'est pas un hasard que, parmi les plus importants représentants du révisionnisme modéré sur la vertu, on retrouve des références fortes à des philosophes du « sens commun » tel que Hume et Wittgenstein.

En tout cas, nous parlons d'abandon d'ambitions, parce que la théorie de la vertu classique a été clairement ambitieuse dans les deux sens. Les classiques de la psychologie de la vertu ont formulé des analyses raffinées de psychologie morale, en ce qui concerne le rôle des émotions et du raisonnement dans la conduite excellente. Toutefois, il y a le problème de la pertinence par rapport à la psychologie empirique contemporaine. En ce sens, le modèle d'Aristote est un représentant excellent de la psychologie des vertus classique. La version aristotélicienne de la théorie de la vertu classique est celle qui maximise la possibilité de dialogue entre psychologie contemporaine et tradition philosophique.

Tout en restant dans le cadre d'une approche unitaire de la vertu (les vertus morales et la prudence ne peuvent être apprises et maîtrisées qu'en même temps) le pluralisme aristotélicien (les vertus morales sont distinctes et peuvent entrer en conflit) favorise le dialogue avec la psychologie

³⁷ Dans la tradition néo-wittgensteinienne (qui n'est pas ambitieuse d'un point de vue psychologique), on s'intéresse pourtant souvent à l'excellence morale. C'est le cas de Hursthouse, *On Virtue Ethics*.

³⁸ L'éthique de la vertu ancienne était critique par rapport à l'éthique de la vertu « populaire », implicite dans le sens commun de l'époque.

contemporaine de l'expertise, qui, comme le verrons dans le chapitre qui lui est consacrée, est complètement axée sur la nécessité d'une spécialisation.

La flexibilité de l'individu vertueux se rapproche également de la conception que nous avons aujourd'hui de la cognition de l'expert. L'individu vertueux a des habitudes et des réactions émotionnelles extrêmement fines et calibrées. Il est aussi en mesure de raisonner sur des nouveaux cas et d'adapter ses réactions habituelles à des compromis et à des arbitrages qui se révèlent appropriés. Le raisonnement a comme base une sensibilité raffinée, cette sensibilité est flexible et suit les conseils de la raison. La psychologie contemporaine nous donne une image très proche de la cognition de l'expert, qui a en même temps les habitudes les plus raffinées et les plus flexibles. Même s'il s'agissait de suivre mécaniquement ses propres réflexes, l'expert l'emporterait facilement sur le simple connaisseur. Mais l'expert a aussi la plus grande capacité à adapter ses réactions automatiques. La cognition automatique de l'expert devient intelligente³⁹.

³⁹ Voir dernier paragraphe du chapitre sur l'expertise et M. Bilalic, P. McLeod et F. Gobet, "Inflexibility of experts—Reality or myth? Quantifying the Einstellung effect in chess masters," *Cognitive Psychology* 56, no. 2 (2008).

Introduction à la première partie

La notion de personnalité et de caractère sont aussi familières que vagues. En un sens, toute dimension de variation individuelle pertinente pour le fonctionnement social et l'épanouissement de l'individu relève de la personnalité ou du caractère. Ainsi, il y a des personnes qui sont courageuses, d'autres qui sont déterminées ou timides : il s'agit des traits de caractère. Mais une personne peut correspondre aussi à un profil défini par une classification typologique spécifique, générale ou relevant d'un domaine particulier. Il y a des hyperactifs, des sur-contrôlés, des dépressifs (général), ou alors des technophiles, des étudiants distraits, des clients passifs (spécifique). Les personnes sont également différenciées par l'image d'elles-mêmes qu'elles entretiennent, par l'attitude qu'elles ont face à leurs valeurs, par la présence de buts et de projets sur lesquels leurs vies sont bâties, par la présence d'une identité fixe ou plus contextuelle et fluide. Ces processus relèvent de ce qui est appelé « le soi ».

Nous avons énuméré des traits, des types, des valeurs, des buts ou des projets qui sont enracinés dans une série de processus d'individuation. La liste pourrait être facilement prolongée. D'un côté, il y a toutes sortes de traits (généraux ou spécifiques, ex : être généreux en général ou compréhensif dans le milieu du travail), de typologies, de valeurs et de processus d'individuation. De l'autre côté, toute structure ou tout processus psychologiques capables d'affecter la variation individuelle dans les attitudes sociales et l'épanouissement personnel devraient être pris en compte. Ainsi, il y a autant d'entités psychologiques qui relèvent de la personnalité que de théories qui étudient la cognition sociale et le développement de l'individu. La dimension de la personnalité est vaste, plurielle, et ses limites sont mal définies.

Les mots « personnalité » et « caractère » ne sont pas synonymes, mais aucune détermination de leur différence n'est universellement acceptée, ni en philosophie ni en psychologie, ou selon le sens commun. La tendance à distinguer entre au moins deux sens ou deux niveaux d'analyse témoigne de l'exigence de mettre de l'ordre dans le pot-pourri des notions qui relèvent de la personnalité. Sans un travail de clarification conceptuelle qui fasse un peu d'ordre, on ne risque pas uniquement d'aboutir à des *mauvaises* théories de la personnalité, on risque de mettre à mal la possibilité d'une *science* de la personnalité, c'est-à-dire d'un cadre conceptuel conceptuellement rigoureux et empiriquement solide qui puisse être accepté par toutes les approches qui s'intéressent à des aspects plus spécifiques de la variation individuelle.

Etant donné que le but de cette thèse est d'instaurer un dialogue entre théorie de la vertu et psychologie empirique, il faut constater que la quantité de préliminaires nécessaires sera très importante. D'abord, pour qu'il y ait un dialogue, il faut qu'il existe une science de la personnalité et que son contenu soit pertinent par rapport aux intérêts traditionnels des philosophes. Cette pertinence doit être démontrée aussi bien en principe que de façon substantielle. En principe, car, dans le cadre d'une démarche normative de *transformation* du caractère, l'intérêt de considérations factuelles sur la structure de la personnalité n'est pas évident. Il faudra aborder la question épineuse du lien entre être et devoir être en philosophie morale et psychologie philosophique : ce sera l'objet du premier chapitre. De façon substantielle, car les connaissances déterminées sur la personnalité dont nous disposons à l'heure actuelle pourraient ne pas mettre en question les thèses des philosophes, même si un dialogue était conceptuellement possible : la deuxième partie de la thèse sera consacrée à l'exposition des théories scientifiques de la personnalité et à la démonstration de leur pertinence philosophique.

Après avoir prouvé (en principe) la pertinence des données factuelles en philosophie et avant d'analyser des théories spécifiques, il faut montrer qu'une science de la personnalité sérieuse existe. Une véritable science de la personnalité doit résoudre ses problèmes de filiation problématique avec le sens commun : toutes les notions qui relèvent de la personnalité sont empruntées ou au moins inspirées du sens commun. L'usage ordinaire des concepts de trait ou de type est trop polysémique, ambigu et contextuel pour que la science puisse les mobiliser sans un processus de désambiguïsation et de nettoyage conceptuel préalable. Le troisième chapitre sera consacré à l'étude de l'usage des termes de trait par le sens commun. À son tour, cette étape demande une préparation : dans le deuxième chapitre, nous analyserons les différentes approches pour déceler et interpréter les théories implicites dans la cognition ordinaire.

Enfin, dans le dernier chapitre de la première partie nous essaierons de mettre de l'ordre dans la pluralité des niveaux d'analyse de la personnalité. Il faut reconnaître que le foisonnement d'approches théoriques différentes et incompatibles qui caractérise l'histoire de la réflexion sur la personnalité tient de la complexité et du caractère pluridimensionnel du phénomène étudié. En ce sens, face à une réalité complexe et stratifiée, tant qu'on n'a pas trouvé un cadre conceptuel adéquat, il est impossible d'éviter le réductionnisme. Sans le soutien d'un cadre conceptuel fort, les théories qui démultiplient les entités et les processus paraissent *ad hoc* et sans véritable pouvoir explicatif. Ainsi, parmi les entités et les processus dont le rôle dans le façonnement de la personnalité est plausible (tempérament, traits, types, valeurs, buts, projets, défis personnels, émotions, sentiments, croyances, désirs, etc.), à chaque fois les théories de la personnalité ont fait un choix, cherchant à maximiser le rôle d'un sous-ensemble et à minimiser le rôle joué par le reste.

Cette situation a contribué à la perte de crédibilité de la psychologie de la personnalité : des théories *centrées* sur les traits se sont opposées à des théories centrées sur les valeurs, les buts ou les projets de vie censés définir et stabiliser la personnalité de l'individu. Aucune approche ne semble avoir les moyens de l'emporter et il est difficile de ne pas voir de l'arbitraire dans ces débats.

Le cadre conceptuel nécessaire pour mettre en perspective les différentes approches spécifiques nous est offert par la théorie de l'évolution. L'hypothèse d'une architecture cognitive « stratifiée » typiquement évolutionniste s'accorde très bien avec nos connaissances empiriques sur le fonctionnement de la personnalité. Souvent l'évolution rajoute des systèmes plus récents à des systèmes plus anciens dans la phylogenèse, tout en *préservant* en partie le mode de fonctionnement des derniers : le résultat est un système stratifié où plusieurs couches interagissent. La théorie de l'évolution paraît en mesure d'articuler un cadre non-réductionniste et non *ad hoc* de l'architecture cognitive complexe de la personnalité humaine. Les détails de ce cadre sont forcément spéculatifs (quoique cohérents avec nos meilleures connaissances), et il faut surtout retenir le projet d'une approche non-réductionniste motivée par la compréhension de l'architecture évolutionniste de la personnalité.

À partir de l'ensemble de nos connaissances sur le fonctionnement de la personnalité, nous pouvons proposer une architecture cognitive à quatre niveaux, chacun ayant probablement une histoire évolutive spécifique :

le tempérament : un ensemble de dispositions générales mais peu fiables, avec une base neurologique claire et des homologues chez les animaux. Par exemple, des tendances générales à l'extraversion ou au névrotisme sont générées par des systèmes neurobiologiques (respectivement) de motivation et d'inhibition, très anciens et bien étudiés chez l'homme.

les stratégies comportementales ou traits contextuels : peut-être aussi chez l'animal, mais certainement chez l'homme, il existe des traits plus spécifiques et contextuels, mais plus fiables. Par exemple, une timidité ou une anxiété sociale qui visent une liste spécifique de situations, et pas d'autres. Dans la situation concernée, l'attitude se manifeste de façon très fiable.

les outils de gestion de la réputation : comme nous montrerons dans le chapitre sur le sens commun, il existe dans toutes les cultures une série de concepts évaluatifs (trait, type) mobilisés dans le processus de définition et redéfinition de la réputation, dans un contexte social. Ce processus est source de variation individuelle, car la stabilisation et l'intériorisation d'une réputation affectent systématiquement les attitudes comportementales.

le « soi » : enfin, tous ces niveaux stratifiés de variation individuelle sont mobilisés et intégrés dans un processus d'individuation qui aboutit à la définition d'une identité personnelle.

D'après tout standard raisonnable, notre connaissance scientifique de la personnalité en est à ses balbutiements. Le rôle de ce cadre conceptuel est surtout de montrer que la personnalité est un phénomène psychologique qui a une identité claire et une histoire évolutive spécifique. Il s'agit d'une « espèce naturelle psychologique », c'est-à-dire un domaine sur lequel des généralisations scientifiques intéressantes et non contingentes peuvent être formulées. Dans la deuxième partie nous allons procéder à la démonstration de la pertinence pour la théorie de la vertu des théories psychologiques sur la personnalité. La troisième partie sera consacrée à l'évaluation des ambitions de la théorie de la vertu à la lumière des approches psychologiques pertinentes.

La nécessité d'une psychologie des vertus basée sur la psychologie empirique

La théorie des vertus est une théorie normative de la formation et du développement continu du caractère moral. Nous allons défendre l'idée selon laquelle la psychologie normative des vertus doit être bâtie à partir de la compréhension de la structure de la personnalité qui émerge en psychologie empirique. L'intuition est la suivante : le point de départ de toute théorie normative du caractère – aussi ambitieuse soit-elle – doit être la prise en compte de la psychologie de la personnalité réelle des gens. Par « prendre en compte », j'entends « ne pas pouvoir ignorer – comme si cela n'existait pas – pour justifier ses conclusions » : cela est tout à fait compatible avec la volonté de changer les modes de fonctionnement psychologiques qu'on a découverts.

Face à des choix méthodologiques aussi basiques – exiger ou ne pas exiger d'une théorie normative la prise en compte des découvertes de la psychologie expérimentale – il est fréquent d'arriver à une impasse dialectique. D'un côté, on a l'impression de prêcher à des convertis : en sciences humaines, toute une série de « sensibilités théoriques » (de la psychologie empirique aux sciences cognitive, en passant par tous les différents projets de naturalisation en sciences sociales) donnent pratiquement pour acquis le rôle d'une étude empirique de la psychologie humaine. Sans cela, paraît-il, si le projet est l'amélioration de la condition humaine, on ne saisirait même pas ce de quoi on parle. De l'autre côté, dans la tradition de la philosophie normative de la rationalité et de la subjectivité du XX^{ème} siècle, on ne voit pas pourquoi des découvertes peu flatteuses sur les limites du fonctionnement psychologique standard des gens mettraient à mal l'importance d'idéaux normatifs ambitieux. Pourquoi les biais du jugement commun, l'usage d'heuristiques approximatives et toutes sortes d'influences situationnelles inattendues devraient-ils nous pousser à abandonner des idéaux ambitieux d'autonomie et rationalité ? S'agit-il d'un manque d'ambition normative, sous prétexte d'adéquation empirique ? Un idéal éloigné et difficile est-il un idéal à abandonner ? La philosophie a souvent incarné un projet radical d'amélioration de la condition humaine : la difficulté ne devrait qu'en souligner l'urgence. Seul un conservatisme masqué de réalisme pourrait s'opposer au projet philosophique.

Dans une impasse dialectique de ce genre, on risque de se faire écouter aussi peu par ses alliés que par ses opposants. Les premiers, aussi bien que les seconds, sont déjà convaincus à l'avance que les détails d'une argumentation méthodologique ne sont pas significatifs. Pour éviter l'impasse, il faut avant tout montrer que la question ne saurait trouver une solution immédiate et évidente. Il y a des

arguments intéressants d'un côté comme de l'autre, et les raisons qui sont fréquemment mobilisées pour liquider rapidement l'une des deux approches ne sont pas convaincantes. Néanmoins, nous chercherons à montrer que l'idée d'un partage net entre d'un côté une psychologie de la vertu normative (justifiée d'une manière satisfaisante sans mobiliser massivement les découvertes de la psychologie empirique) et, de l'autre, une pratique de l'éducation du caractère des personnes réelles (à laquelle on laisserait toute la responsabilité de gérer les limites psychologiques dans le processus d'implémentation des structures normatives) n'est pas plausible. La psychologie empirique du caractère est pertinente à tous les niveaux, normatif *et* pratique. Cependant, la thèse contraire n'est ni immédiatement absurde ni conceptuellement incohérente. L'idée d'un projet normatif justifié de façon indépendante des détails d'implémentation a ses mérites et ne manque pas de défenseurs sophistiqués, y compris dans le cas du caractère. Toutefois, en dernière instance, nous croyons qu'elle n'est pas défendable. La raison principale en est la suivante : à moins que l'on adopte une méta-éthique et une théorie de la normativité explicitement conçues pour délivrer le résultat voulu, il n'est pas plausible qu'un idéal de fonctionnement et de structuration du caractère humain puisse être justifié *indépendamment* de la matrice des coûts, des opportunités et des bénéfices nécessaires pour atteindre un tel résultat, étant donné la psychologie réelle des gens.

L'insuffisance du débat contemporain sur l'adéquation empirique de la théorie de la vertu face à la menace situationniste : la rhétorique de « la philosophie qui nie l'évidence factuelle »

Avant d'approfondir l'histoire et la structure du problème de l'adéquation empirique d'une théorie normative du caractère, une référence au contexte où cette question est discutée dans le débat philosophique actuel de langue anglaise est inévitable, moins pour situer le problème que pour se démarquer d'une approche de la question que nous considérons insatisfaisante. À l'heure actuelle, la question est débattue uniquement dans le contexte de la réponse que les théoriciens de la vertu organisent face à la menace situationniste radicale, avancée par des auteurs tels que Harman et Doris. On consacrerait un chapitre entier pour articuler et discuter les arguments des situationnistes, ici il s'agit de montrer à quel point il est insatisfaisant de réduire le débat général sur l'adéquation empirique à « comment répondre aux situationnistes ».

Au niveau formel, les situationnistes accusent les théoriciens de la vertu de mobiliser une notion de trait qui n'a pas de référence dans la réalité. La psychologie nous montrerait que le type de traits associés aux vertus n'existe pas. La forme est celle d'une objection *spécifique et interne*. Spécifique, parce que la critique ne s'inspire pas d'une évaluation globale de la stratégie des philosophes. Au contraire, elle vise un problème déterminé : la mobilisation d'une notion

inadéquate de trait. Interne, parce qu'elle n'a pas besoin de prémisses indépendantes pour être valable, mais déniche un problème évident dans la démarche de la théorie de la vertu. En d'autres termes, le situationniste ne dit pas au théoricien de la vertu : « je sais comment tes affaires devraient être gérées globalement, et tu t'y es trompé du début à la fin » mais plutôt « fais ce que tu veux en général, mais tu n'as pas le droit de présupposer cela ».

Les objections spécifiques et internes sont normalement plus puissantes et décisives, pour une raison claire : elle demandent moins de présupposés pour fonctionner. Face à un problème évident, le discours est disqualifié quoi qu'il arrive, alors que si un discours est incompatible avec un arrière-plan complexe de considérations, la force de la critique dépend de la justification de cet arrière-plan. La volonté de développer des objections spécifiques et interne est donc bien compréhensible. Cependant, dans le contexte de ce débat, cette façon de procéder paraît insatisfaisante pour au moins trois raisons : (1) il s'agit d'un effet rhétorique, car, si on regarde les détails, la nature profonde des arguments situationnistes n'est pas celle d'objections spécifique et internes ; (2) elle est en même temps trop (dans l'interprétation des données psychologiques substantielles) et trop peu (au niveau méthodologique) radicale contre la théorie de la vertu traditionnelle ; (3) elle a complètement dénaturé le débat : la plupart des philosophes ont l'impression de devoir répondre à un problème très spécifique, posé par une tradition psychologique minoritaire qui mobilise trois ou quatre paradigmes expérimentaux, comme s'il agissait de réparer un pneu d'une voiture qui par ailleurs marche.

(1) En ce qui concerne le premier point, comme ce sera clair après l'analyse détaillée des arguments situationnistes dans le chapitre qui y est consacré, celle d'une objection interne et spécifique n'est qu'une façade rhétorique. Le choix de cette stratégie rhétorique de la part de Harman et Doris (surtout Doris, qui a développé davantage la discussion et a imposé cette formulation) dans le meilleur des cas n'a fait que créer de la confusion sur des arguments intéressants et, dans le pire scénario, a été utile pour masquer les faiblesses des arguments. Superficiellement, l'accusation est uniquement celle d'endosser des prémisses empiriques inadéquates, au sens technique d'accepter de faux présupposés de nature factuelle. Doris insiste sur le fait qu'une théorie éthique ne puisse pas avoir des prémisses empiriques, aussi cachées soient-elles. Cela ferait des théories éthiques la cible potentielle facile pour des objections internes puissantes et inexorables, qui se limiteraient à montrer ou faire remarquer que les prémisses concernées sont fausses. Propre et efficace, car aucun philosophe ne peut nier que *si* une théorie dépend effectivement de prémisses factuelles fausses, cela représente un problème. On a l'impression que des considérations méthodologiques complexes sont stratégiquement inutiles, étant donné la présence d'un point faible aussi facile. Mieux vaut

laisser la liberté à une théorie éthique de se développer comme elle veut, pourvu que ses prémisses factuelles soient défendables.

Sauf que les choses sont plus complexes : aucune théorie n'expose ses présupposés empiriques d'une façon claire et non contestable. Est-ce que la théorie de la vertu présuppose l'existence des traits dans la population ? Ce n'est pas clair : ce que la théorie affirme est que des traits peuvent être développés, non pas qu'ils sont répandus. Doris est bien conscient de cela, et finalement son vrai argument se révèle être plus sophistiqué : une longue *évaluation globale* de la littérature empirique montre que développer des vertus serait très difficile, si bien qu'*au bout du compte*, il n'est pas raisonnable de la part d'une théorie éthique de nous demander cela. Bref, l'argument en réalité consiste en une évaluation méthodologique complexe sur la façon dont une théorie normative du caractère devrait être construite : rien à voir avec le fait de remarquer qu'un théoricien de la vertu n'a pas le droit de se tromper sur le nombre des lunes de Jupiter !

Sur la substance de cette remarque, Doris est certainement d'accord. Mais alors pourquoi ne pas présenter directement l'attaque situationniste comme « un ensemble de considérations complexes sur les liens inextricables entre l'empirique et le normatif dans la structure d'une théorie de la vertu », au lieu de répandre l'image de philosophes qui nieraient sans vergogne l'évidence factuelle ? La rhétorique situationniste qui tente de ridiculiser la position des philosophes n'est pas complètement innocente. Car l'objectif du situationniste radical est celui de liquider l'idée d'une théorie normative du caractère : liquider les théories traditionnelles sous prétexte qu'elles nient les faits est un meilleur moyen pour tourner la page que s'embarquer dans une analyse des enjeux méthodologiques et des *desiderata* d'une théorie normative du caractère. La deuxième approche invite à explorer davantage de solutions alternatives pour satisfaire les ambitions de la tradition, ce que nous cherchons à faire dans cette thèse et que Doris refuse, en choisissant d'explorer une éthique situationniste radicale. La rhétorique finit ainsi par nuire à la qualité argumentative substantielle, car, si on analyse une question à l'envers on finit par ne pas être aussi clair qu'il le faudrait et on n'aborde jamais du début à la fin l'ensemble des questions décisives.

(2) En ce qui concerne le deuxième point, la rhétorique de « nier l'évidence » annonce des résultats exagérément catastrophiques pour la confrontation entre les ambitions perfectionnistes de la théorie de la vertu et les données empiriques. On verra que, si la structure du caractère vertueux envisagée par la tradition n'est pas tenable, d'autres conceptions du caractère restent ouvertes à l'exploration empirique. Sur ce plan substantiel, le situationniste est trop radical. Sur le plan méthodologique, par contre, il ne l'est pas suffisamment. Car la rhétorique de l'objection spécifique et interne favorise l'idée selon laquelle – pour peu qu'elle ne nie pas l'évidence – une théorie éthique pourrait procéder comme elle veut. En réalité, l'objet du scandale devrait être le refus de la philosophie morale de

reconnaître l'interpénétration complexe entre éléments empiriques et éléments normatifs dans la méthodologie d'une théorie normative. Cette interpénétration riche et stratifiée ne se réduit pas à « accepter l'évidence », tout comme son refus n'est pas simplement une façon de nier les faits. On ne doit pas se réveiller face à des données problématiques, on aurait dû intégrer les données de la psychologie depuis le début.

(3) Les conséquences les plus graves de la rhétorique situationniste se mesurent aux effets néfastes sur le débat. Réagissant à la menace situationniste, les théoriciens de la vertu n'ont pas compris qu'il s'agit de repenser des aspects profonds de leur philosophie de la psychologie et de leur méthodologie. Ils ont essayé de contrer l'objection spécifique des situationnistes, en montrant que les données empiriques mobilisées n'arrivent pas à réfuter les prémisses empiriques dont il est question, notamment l'existence des traits de caractère. Tout comme s'il s'agissait de vérifier si un pneu d'une voiture – qui par ailleurs marche sans problèmes – est crevé ou non, au lieu de se demander si le projet de la voiture est bien conçu et lui permet de tenir la route. Le problème de fond est méthodologique et ne disparaîtrait pas si les nouvelles provenant des données empiriques étaient moins mauvaises qu'elles ne le sont.

Un deuxième inconvénient généré par l'impression d'être confronté à un problème circonscrit est celui de favoriser des solutions *correctives*, lorsque l'objection paraît suffisamment convaincante pour ne pas être tout simplement rejetée. Si le problème est spécifique, mieux vaut faire quelques petites modifications qui ne touchent pas à la structure générale de la théorie. Il s'agit de la deuxième forme de réaction typique des philosophes, exemplifiée par l'attitude de Kamtekar : *« Loin d'être mis à mal par des découvertes empiriques portant sur les faiblesses pérennes du raisonnement humain, l'idéal aristotélicien de caractère vertueux peut être reformulé prenant en compte les stratégies d'amélioration du raisonnement issues des science empiriques⁴⁰. »*

Partagés entre la perspective d'ignorer ou de résoudre un problème circonscrit, les philosophes n'ont pas pour l'instant pris la mesure du défi lancé par la psychologie.

Le dialogue avec la psychologie empirique est-il possible ? Philosophie de la psychologie wittgensteinienne et données expérimentales

Les théoriciens contemporains de la vertu ont presque systématiquement ignoré le corpus très important de données empiriques accumulées au fil du XXème siècle par la psychologie de la personnalité et la psychologie sociale. L'importance de cette rencontre manquée demande une

⁴⁰ Rachana Kamtekar, "Updating Practical Wisdom," in *Virtue Ethics, Character and Moral Psychology*, édition C. Upton (Oxford: Oxford University Press, à paraître).

explication. Le premier choix interprétatif à faire est le suivant : un simple phénomène d'ignorance réciproque (peut-être explicable au niveau de sociologie de la science) ou une mise à l'écart des données psychologiques pour des raisons profondes au niveau théorique ? Etant donné la rareté de références textuelles dans la littérature philosophique, surtout avant les travaux pionniers de Flanagan au début des années 90, mais aussi après, dans la majorité des textes cruciaux récents sur la psychologie de la vertu, la première option est très tentante⁴¹. Il est difficile de résister à la conclusion que si les philosophes ne connaissaient pas la psychologie de la personnalité c'est à cause d'un manque de compréhension total de son sujet réel.

Cependant, il vaut mieux résister à cette tentation pour deux raisons. D'un côté, nous n'avons pas les ressources nécessaires pour étudier ce phénomène remarquable d'ignorance réciproque avec les moyens et les normes critiques de la sociologie de la science. De l'autre, il est plus constructif d'assumer l'attitude la plus charitable possible : nous voulons donc prendre au sérieux cette absence de références à la littérature empirique, cherchant à en dénicher les raisons implicites profondes. Dans cette démarche, on peut envisager deux options : (1) il est possible que les données empiriques soient tenues à l'écart parce que la philosophie normative du caractère est une entreprise largement indépendante de la psychologie empirique : par principe, pour des raisons méthodologiques de fond ; (2) il se peut qu'il n'y ait aucune barrière méthodologique : il est tout à fait conceptuellement envisageable que des données expérimentales mettent à mal une théorie des vertus, mais pour l'instant les philosophes ne voient aucune menace dans la littérature empirique concernée. En d'autres termes, on pourrait découvrir expérimentalement que l'éthique de la vertu n'est pas tenable, mais les données pour l'instant sont favorables ou neutres.

(2) Cette dernière position (largement implicite, car des prises de position explicites sont très rares) semble être la plus répandue aujourd'hui. À cause de l'attaque explicite à la théorie de la vertu de la part de la psychologie sociale et du climat intellectuel favorable à l'interdisciplinarité, il est de plus en plus fréquent de trouver des exhortations à l'étude empirique des vertus et de moins en moins fréquent de faire face à des réactions de rejet contre la psychologie, qui aurait dépassé les limites de son domaine de compétence. Cette nouvelle attitude méthodologiquement ouverte, mais sceptique quant aux conséquences philosophiques de littératures psychologiques spécifiques, est la plus défendable et celle qui mérite une réponse plus longue. Il faut élaborer une interprétation alternative d'un corpus important de littérature empirique, pour en montrer la pertinence. Plusieurs chapitres de

⁴¹ Owen J. Flanagan, *Varieties of moral personality: ethics and psychological realism* (Cambridge, Mass.: Harvard University Press, 1991). Cet ouvrage a lancé le débat actuel sur l'adéquation empirique de la théorie de la vertu. Des classiques contemporains en éthique de la vertu ressort une ignorance ou une compréhension totalement inadéquate des travaux en psychologie de la personnalité.

cette thèse ne sont qu'une longue réponse à un philosophe convaincu de la nécessité d'un dialogue avec la psychologie, mais qui pour l'instant ne voit pas de menace pour les thèses traditionnelles de la théorie de la vertu.

(1) La première position, quoi qu'en perte de vitesse, a joué un rôle fondamental dans l'histoire de la théorie de la vertu contemporaine, par deux biais indépendants : l'influence lourde de la philosophie de la psychologie de Wittgenstein et une compréhension inadéquate de l'idée d'une psychologie philosophique normative.

Parmi les auteurs avec un arrière-plan wittgensteinien on peut citer les plus grands noms de l'éthique de la vertu néo-aristotélicienne contemporaine : Elizabeth Anscombe, Philippa Foot, Peter Geach, Rosalind Hursthouse, Alasdair MacIntyre, Bernard Williams, John McDowell, Anthony Kenny, Sabina Lovibond et Georg Henrik von Wright. L'influence est aussi importante que mystérieuse, car, comme cela a été remarqué : « *Il semble qu'il n'y a aucun lien nécessaire entre la philosophie de la psychologie de Wittgenstein et l'éthique de la vertu*⁴² ».

Dans le cadre de la philosophie de la psychologie wittgensteinienne (qu'ici on ne pourra pas résumer de façon satisfaisante), on rejette l'idée selon laquelle les explications psychologiques du sens commun sont des explications causales et se réfèrent à des entités qui ont une véritable influence causale sur le comportement. Quiconque procède à une interprétation causale se rend coupable d'une confusion conceptuelle engendrée par une incompréhension réflexive de concepts que pourtant il maîtrise dans la pratique. On rend compte du comportement d'autrui tous les jours en mobilisant des mots tels que « désir », « croyance » ou « intention ». Cette pratique n'a pas besoin de révision substantielle. Pourtant, dès qu'on assume une posture philosophique, on est égaré par des ressemblances superficielles d'usage entre le langage psychologique et le langage causal.

Il y a une ressemblance formelle entre « il cherche sa clé sur la table *parce qu'il* veut sortir et qu'il est convaincu de l'avoir laissée là » et « la voiture se met en marche *parce que* le moteur a été allumé ». Cela nous pousse à considérer qu'il s'agit d'explications causales dans les deux cas. Il y aurait des « choses » comme des « désirs » et des « croyances » qui sont causalement responsables de nos actions, tout comme un moteur est responsable de la mise en marche d'une voiture. D'après Wittgenstein, le statut ontologique des croyances et des désirs n'est pas celui de « mécanismes dans

⁴²Monahan, "Meeting Anscombe's Demand: Toward a Moral Psychology of Character" 19. Cependant, un lien stratégique plutôt évident est le suivant : une approche qui considère que réviser ou abandonner des concepts ordinaires sur la base d'expériences psychologiques relève de la confusion conceptuelle a tout intérêt à exploiter les concepts les plus riches, puissants mais aussi problématiques du sens commun (dont sans doute celui de vertu). Qui ne craint pas l'adéquation empirique a tout l'intérêt à exploiter les concepts empiriquement plus problématiques.

la tête » mais celui de dispositions complexes au comportement : considérer une croyance comme une cause serait une confusion conceptuelle du même type que parler de la couleur des nombres. On n'a pas le droit de considérer les croyances, les désirs et les intentions comme des facteurs causalement actifs pour la raison suivante (entre autres) : contrairement aux états de conscience qui sont vraiment « dans la tête » (sensations, impressions, images mentales), dire qu'on entretient un désir, une croyance ou une intention exactement d'un instant t à un autre instant t' n'a pas de sens. Si on ne peut pas dire qu'une condition psychologique a une durée exacte, il s'agit plutôt d'une disposition générale qui explique le comportement d'un point de vue externe.

D'une personne qui croit qu'il pleut on s'attend à un ensemble de comportements typiques, tout comme on s'attend à une certaine fluidité du trafic dans un croisement régulé de façon intelligente par des feux. Comme il serait absurde de penser qu'il y a « quelque chose » comme une « capacité à gérer le trafic » physiquement active à l'intérieur des feux, il est absurde d'imaginer « quelque chose » comme une croyance causalement active dans la tête des gens. Une croyance n'est pas plus dans la tête que la fluidité du trafic n'est physiquement générée par les feux.

Les traits de caractère sont censés être aussi des dispositions générales, aux côtés de croyances, désirs et intentions. D'après Wittgenstein et les théoriciens de la vertu influencés par sa philosophie de la psychologie, les explications qui mobilisent des traits de caractère ne sont pas des explications causales. Apparemment, cette thèse donne une autonomie énorme à la théorie de la vertu : elle se serait débarrassée d'un seul coup du problème de l'adéquation empirique (car la psychologie empirique est censée s'intéresser uniquement à des processus causaux) et de deux formes de scepticisme qui l'ont souvent visée : le « scepticisme absolu » et le « scepticisme relatif » contre les explications en termes de traits de caractère.

Le scepticisme absolu est la thèse selon laquelle les explications qui mobilisent des traits de caractère sont des pseudo-explications : en réalité, elles n'expliqueraient rien du tout. Dire « il a fait cela parce qu'il est courageux » serait donner une fausse explication (souvent l'accusation est de circularité : l'action concernée est le genre de chose que seul un courageux peut faire). D'après le scepticisme relatif, les explications en termes de traits de caractère n'ajoutent rien à d'autres explications qui mobilisent des concepts plus simples, tels que croyance, désir et intention : par conséquent, il vaudrait mieux utiliser directement ce langage plus clair. Au lieu de dire « il a fait cela parce qu'il est courageux », il faudrait dire « il a cru que l'enfant était en train de se noyer et que personne d'autres ne pouvait l'aider et il a aussitôt décidé de se jeter à l'eau ».

Ces deux formes de scepticisme se retrouvent très régulièrement parmi les critiques adressées à la théorie de la vertu contemporaine et circulent très largement dans les milieux intellectuels où la théorie de la vertu n'est pas appréciée. L'adoption d'une philosophie de la psychologie

wittgensteinienne neutralise le problème du scepticisme absolu et celui de l'adéquation empirique du même coup : l'attribution de traits de caractères n'est pas censée offrir une explication causale mais uniquement repérer un schéma intéressant dans le comportement. Dans cette perspective, il faudrait concevoir (par exemple) le courage davantage comme un concept de statut, tel que « niveau socio-économique ». Le niveau socio-économique d'une personne est une mesure qui résume une série de dimensions plus basiques et plus concrètes telles que salaire, niveau d'instruction, quartier de résidence, etc. Cela permet de générer des prédictions sur les choix et les attitudes dans la vie quotidienne. Attribuer un niveau socio-économique élevé est croire qu'un ensemble complexe de dynamiques est en place, mais il ne s'agit pas de postuler un facteur causalement actif dans la tête des gens. Le niveau socio-économique ne conditionne pas le comportement de la même façon que la disponibilité d'un système visuel qui marche correctement. Les gens n'ont pas littéralement un « système » socio-économique. Il s'agit plutôt d'une mesure commode pour avoir une idée de l'effet conjoint de plusieurs facteurs plus basiques qui eux sont directement actifs.

Attribuer le courage serait repérer une dynamique complexe, au sens suivant : face à une personne courageuse, on s'attend à ce qu'une série de phénomènes typiques puissent se produire. Mais le courage n'est pas quelque chose que le courageux possède littéralement. De même pour les autres traits : lorsque Marc est dans une soirée, tout le monde rigole et s'amuse, car il est très sympathique. Cependant, d'après le wittgensteinnien, la sympathie n'est pas une propriété psychologique réellement active dans la tête de Marc : il s'agit d'une dynamique « écologique » complexe qui se déploie bien en dehors de son crâne. La sympathie est tout autant dans la salle où a lieu la soirée que dans la tête de Marc. Ces *patterns* de comportements n'offrent pas de véritables explications : pour preuve, ils ont bien évidemment besoin d'être expliqués eux-mêmes. Qu'est-ce qui rend possible un comportement courageux stable, la sympathie, la générosité, un niveau socio-économique élevé ? Une explication qui mobilise d'autres concepts est nécessaire.

À ce stade de la dialectique, il semble que le théoricien de la vertu néo-wittgensteinien soit mal placé pour répondre au scepticisme relatif : si les traits de caractère n'expliquent rien, pourquoi ne pas essayer avec une explication en termes de croyances, désirs, intentions ? Ici ressort l'originalité de la position wittgensteinienne, par rapport à un scepticisme beaucoup plus répandu soulignant la circularité des explications par les traits. Comme on l'a vu, le même statut de « dispositions générales au comportement » est étendu aussi aux concepts de croyance, désir et intention, si bien qu'ils ne sont pas non plus de bons candidats pour une explication alternative. L'interprétation wittgensteinienne des concepts psychologiques du sens commun a un effet très intéressant de préservation contre toute critique, car il élimine la motivation essentielle du processus de révision :

l'adéquation causale. Dans la construction d'une théorie explicative du comportement, il est normal de maximiser la simplicité et le pouvoir explicatif : il faut se demander si un ensemble très vaste d'explications superficiellement différentes peut être réduit à des explications qui mobilisent des concepts plus simples. Ainsi, le contexte d'un projet de théorie causale explique la passion de la philosophie analytique pour le modèle croyances-désirs et la volonté de réduire les traits à des ensembles complexes de croyances et désirs.

L'attitude du wittgensteinien est complètement différente : si les traits de caractère, aussi bien que les croyances et les désirs, sont des dispositions générales au comportement, il n'y a aucune raison pour une réduction réciproque. Ils dénotent des schémas de comportement et des dynamiques complexes différentes : la réduction de l'action de la générosité à l'effet conjoint d'un ensemble de croyances et de désirs n'a pas de sens ; pas plus que la notion de niveau socio-économique ne pourrait être remplacée par un indice général de religiosité. Certes, il peut bien y avoir une corrélation très forte entre un niveau socio-économique bas et une religiosité intense : cela signifie que les facteurs plus basiques (et causalement actifs) dont l'action est résumée par les deux mesures coïncident en partie. Cependant, la réduction complète d'une entité théorique à une autre (habituelle dans le contexte d'explications causales) n'a pas de sens ici. Si on cherche les causes, il est tout à fait raisonnable de se demander si l'action d'un trait n'est rien d'autre que l'action de désirs et croyances déterminés, si bien que, à la rigueur – des raisons pragmatiques d'économie cognitive mises à part – on pourrait arrêter d'utiliser le concept de trait. Mais ce serait une erreur conceptuelle d'identifier deux dispositions émergentes : comme la religiosité ne peut littéralement *être* le niveau socio-économique, un trait de caractère ne peut *être* un ensemble de croyances et désirs (même si toutes sortes de corrélations entre le fait de croire, désirer quelque chose et avoir certains traits sont plausibles)⁴³.

Il y a donc un effet très intéressant de préservation de tout risque de réduction : le seul danger qui reste est celui du simple abandon d'une notion qui n'a plus d'utilité pour rendre compte d'un phénomène. Il est conceptuellement envisageable que dans certaines sociétés (réelles ou possibles) le niveau socio-économique ne soit plus une mesure intéressante, liée à des effets remarquables : dans ce cas, les gens n'arrêteraient pas littéralement d'avoir un niveau socio-économique, mais on

⁴³ En réalité, ce n'est pas techniquement impossible. Laissant de côté la réduction causale, il reste la possibilité d'une réduction par analyse conceptuelle. Tout comme on pourrait éliminer le concept de « célibataire » en le substituant par « non marié », on pourrait éliminer le concept de « trait » en le substituant par « ensemble x, y de croyances et désirs ». Cependant, après des décennies de tentatives, l'analyse conceptuelle des notions psychologiques du sens commun paraît un programme de recherche voué à l'échec. Sur ce point, voir T. Williamson, *Knowledge and Its Limits* (Oxford University Press, 2000).

abandonnerait tout simplement la notion. Pour les traits de caractère, ce risque paraît franchement lointain : si ce n'est par le biais d'une critique aux théories psychologiques (considérées comme étant) causales du sens commun par des théories causales supérieures, ce serait difficile de motiver l'abandon de termes « épais » comme « courage », « sympathie », « générosité », « sagesse ». Cela engendrerait une perte considérable de pouvoir expressif ; il suffit d'imaginer un écrivain qui s'abstiendrait de l'usage de ces mots.

Globalement, le mystère du lien entre la philosophie de la psychologie wittgensteinienne et la théorie de la vertu commence à se dissiper : si, à la base, on est déjà wittgensteinien, les concepts riches et puissants de « trait de caractère » et « vertu » deviennent très intéressants : on peut hériter de la puissance d'un réseau conceptuel, sans aucun désavantage. Les deux problèmes classiques d'une théorie riche qui paraît avoir un contenu empirique important – notamment qu'elle s'expose davantage (1) à des critiques d'adéquation empirique et (2) au risque de réduction en faveur de théories plus simples, sont brillamment écartés. La citation suivante est très significative pour comprendre à quel niveau les théoriciens de la vertu contemporains sont conscients de cette situation stratégique, ils connaissent bien les critiques de leurs adversaires et jouent sur l'avantage d'une position wittgensteinienne :

« Dans la philosophie de l'action contemporaine, il y a un débat fervent pour savoir si chaque action intentionnelle doit être initiée en partie par un désir, ou s'il est possible d'être poussé à l'action par une croyance (...) une prémisse d'arrière-plan du débat dans sa totalité est l'idée selon laquelle les croyances et les désirs sont aussi différents que l'or et l'oxygène, on suppose aussi que la distinction entre le rationnel et le non-rationnel est nette (...)

Pour tous ceux qui sympathisent avec les écrits du dernier Wittgenstein (...) le rejet de distinctions trop nettes en psychologie philosophique est aussi naturel et nécessaire que l'acte de respirer...Je me suis souvent rendue compte du fait que, lorsque j'enseigne l'éthique de la vertu aux étudiants, ou lors de la discussion d'un article (...) avec d'autres philosophes (...) , ce qui bloque la compréhension est la présupposition inconsciente selon laquelle (...), par exemple, croyances et désirs seraient des espèces naturelles, ou qu'une raison [d'agir] serait un couple croyance-désir qui cause une action, ou que les états mentaux seraient des états cérébraux – ou, plus généralement, que la philosophie devrait découvrir ou construire les fondements de notre pensée. S'efforçant de rendre compatibles ces présupposés avec ce qui est dit [l'éthique de la vertu], le public le trouve délibérément obscur ou volontairement incomplet, ou incohérent, ou alors ouvert à des objections si clairement évidentes qu'ils pensent qu'ils ne peuvent pas avoir compris. Quelque fois – pas toujours, naturellement – le

nuage se dissout si on dit "Mais on ne croit pas que les choses soient ainsi si on est wittgensteinien"⁴⁴ ».

Il ne sera pas question ici d'évaluer la philosophie de la psychologie wittgensteinienne, mais de montrer qu'elle n'offre pas la possibilité d'ignorer les données empiriques, comme on pourrait le croire. Il y a au moins deux failles dans la défense wittgensteinienne de l'autonomie de la psychologie philosophique.

(1) La première concerne une analyse correcte du sens commun : dans la perspective wittgensteinienne, le but de la philosophie est d'analyser l'usage des concepts du sens commun pour éviter les confusions conceptuelles générées par une mauvaise compréhension réflexive de ces mêmes concepts. Pour que cette démarche puisse avoir une utilité quelconque pour un théoricien de la vertu, il faut que les résultats montrent l'existence d'une théorie du caractère implicite dans le sens commun. Or, le philosophe, sans l'aide des données empiriques de psychologie sociale et d'anthropologie cognitive, ne peut pas analyser correctement l'usage des concepts par le sens commun. Nous consacrerons un chapitre entier de cette thèse à la conception du caractère implicite dans le sens commun : ici on peut anticiper que, au-delà des apparences et de la sensation de familiarité que chacun d'entre nous ressent envers les notions de caractère et de personnalité, il est très difficile de reconstruire exactement notre théorie implicite. Difficile par rapport à nos meilleures connaissances, sans doute impossible sans prendre en compte les données expérimentales sur l'usage et l'attribution des traits.

Une issue possible consisterait à postuler une spécialisation dans l'usage des concepts du sens commun : des expressions mobilisant des traits sont dans la bouche de tout le monde, mais il faut être un expert pour comprendre la personnalité de quelqu'un. Seulement une minorité d'experts aurait vraiment la maîtrise des concepts de traits de caractère, dont bien sûr les philosophes, qui sont ainsi en mesure de les analyser. Cette stratégie sophistiquée n'offre pas non plus d'immunité face aux données expérimentales : car, pour que les concepts concernés aient une utilité quelconque dans la vie quotidienne, il faut que le groupe d'experts qui les maîtrisent ne se réduise pas aux philosophes. Ce n'est donc pas possible de considérer le philosophe comme la seule autorité dans l'analyse de ces concepts : l'usage qualifié de ceux qui maîtrisent vraiment les concepts de trait peut et doit être étudié expérimentalement.

(2) La deuxième faille concerne la relation entre des découvertes empiriques possibles et l'interprétation des traits comme *patterns* : dispositions générales au comportement qui rendent

⁴⁴ Hursthouse, *On Virtue Ethics* 16.

compte et donnent du sens à ce que font les gens, sans pour autant l'expliquer. Ce n'est pas parce qu'on ne cherche pas à expliquer que les faits réels (tels qu'ils peuvent être avérés par des études psychologiques) ne posent pas des problèmes. D'après le wittgensteinien, en attribuant un trait on reconnaît un pattern dispositionnel. On voit que quelqu'un aide souvent autrui et réagit avec disponibilité et gentillesse aux besoins des gens qu'il fréquente : on peut dire qu'il est généreux. Ensuite, si on découvre que cette personne vole souvent de l'argent à ses amis, il n'y aura aucune explication causale à rejeter, mais on sera au moins obligé d'admettre que l'interprétation initiale n'était pas adéquate. Pour qu'une interprétation soit adéquate, il faut qu'un ensemble de faits se produise. Imaginons de découvrir que personne n'est généreux de façon cohérente dans tous les contextes où une action généreuse serait pertinente et possible. Cela signifie au moins que, dans sa signification originale, le concept de générosité ne peut être appliqué.

Aucune explication causale n'a été avancée ou retirée, mais on ne peut plus rendre compte et interpréter le comportement d'autrui grâce au concept de générosité. Etant donné que, dans cette hypothèse, les conditions requises pour une attribution de générosité sont extrêmement rares, on ne peut rendre compte de l'excellence morale, et surtout de l'*approximation* à l'excellence morale (qui concerne un nombre plus important de personnes) qu'on rencontre dans la vie réelle et dans le témoignage historique. On doit conclure que les traits concernés ne sont d'aucune utilité dans un compte-rendu de la vie morale.

La défense wittgensteinienne de l'autonomie d'une psychologie philosophique normative paraît totalement inefficace. Le genre de découverte empirique qui peut menacer une approche causale à l'explication du caractère est exactement le même qui peut menacer l'approche interprétative. Si on découvre un manque de cohérence comportementale dans la population, il s'en suit que le facteur causal (le trait) censé produire cette cohérence attendue (mais pas trouvée) ne peut exister. Mais il s'en suit également qu'une interprétation qui chercherait à rendre compte du comportement mobilisant une disposition générale (le trait) non explicative ne peut être adéquate. Globalement, en ce qui concerne le problème de l'adéquation empirique et la volonté de rendre autonome la philosophie, être wittgensteinien n'aide en rien.

Psychologie philosophique normative

La philosophie de la psychologie wittgensteinienne se présente comme une stratégie sophistiquée et ingénieuse pour isoler la philosophie des résultats de la psychologie empirique sur la structure de la personnalité. Si on laisse de côté cette option, il semble que la voie pour un dialogue avec la psychologie empirique devrait être grande ouverte. Il n'en est pas ainsi. Paradoxalement, en ce qui

concerne l'adéquation empirique, la situation ne fait qu'empirer chez les non-wittgensteiniens, à cause du modèle de psychologie philosophique normative qui s'est imposé dans le reste de la scène philosophique.

Une grande partie du XXème siècle a été dominée par l'idée selon laquelle on pourrait construire des modèles psychologiques normatifs « causaux » indépendamment de la psychologie des gens telle qu'elle est étudiée par les travaux empiriques. Pour éviter des confusions, ces modèles sont causaux au sens où ils postulent l'action de certaines dynamiques causales dans la psychologie d'un agent *idéal*, non pas au sens où ils décriraient la psychologie d'un individu moyen. Mais pourquoi l'approche causale empire la situation par rapport à celle interprétative ? Le théoricien de la vertu wittgensteinien pense hériter d'un corpus de connaissances du sens commun sur le caractère et la personnalité qu'il faut essentiellement analyser et préserver. L'approche causale est plus révisionniste par rapport à la conception du sens commun que la conception herméneutique. Le théoricien causal doit être crédible : il doit fournir une « *concept proof* ». Il est comme un ingénieur futuriste qui envisage un modèle spécifique de vaisseau spatial et doit prouver que – à part une série de questions techniques qui sont bien sûr encore à régler – c'est théoriquement faisable.

Dans cette démarche, le caractère vague des concepts du sens commun – point de départ de toute réflexion théorique sur le caractère – pose un problème. La vertu est associée à des bénéfices et à des performances très ambitieuses sans qu'on puisse avoir de précisions sur son mode de fonctionnement psychologique (non pas dans la réalité, mais dans le modèle idéal du caractère vertueux proposé). Contrairement au wittgensteinien, le théoricien causal est assoiffé de détails sur la psychologie de l'individu vertueux. Cependant, les détails vont être cherchés du côté des intuitions normatives sur le fonctionnement *souhaitable* d'un agent moral et rationnel, et non du côté de la psychologie empirique. Par conséquent, on ajoute souvent toute une couche de détails qui ne fait qu'amplifier et démultiplier le problème d'adéquation empirique de départ, si bien qu'on s'éloigne de la psychologie réelle des gens encore plus et sur davantage de fronts.

Dans l'article très célèbre de Anscombe qui a stimulé le renouveau de l'éthique de la vertu, on retrouve une exhortation très citée à approfondir l'étude de la psychologie des vertus :

« Il n'est pas profitable pour nous à présent de faire de la psychologie morale, cela devrait être laissé de côté au moins jusqu'à ce qu'on dispose d'une philosophie de la psychologie adéquate, ce qui nous manque largement (...) Dans la philosophie d'aujourd'hui il faut une explication de pourquoi un homme injuste est un homme mauvais, ou une action injuste est une action mauvaise ; donner cette explication relève de l'éthique ; mais on ne pourra même pas commencer tant qu'on ne sera pas équipé d'une philosophie de la psychologie adéquate. Car la preuve qu'un homme injuste est un homme mauvais requiert une explication

substantielle de la justice en tant que vertu. Ce domaine de l'éthique nous échappera complètement tant qu'on n'aura pas une théorie de la vertu – un problème qui ne relève pas de l'éthique mais de l'analyse conceptuelle – et de sa relation avec les actions dans lesquelles elle s'exprime : une question, je pense, qu'Aristote n'a pas réussi à éclairer réellement⁴⁵.»

Cette exhortation est toujours citée pour montrer que, déjà au début de l'éthique de la vertu, il y avait la conscience de ce qu'il fallait faire et qui finalement n'a pas été fait. Car, à cause de la très forte influence wittgensteinienne, la production théorique nécessaire pour reformuler les thèses vagues du sens commun a été limitée au minimum possible. Avec quelques exceptions – notamment les théoriciens qui se sont engagés de façon cohérente dans le projet d'une théorie normative causale des vertus, par exemple Dent et Mohanan⁴⁶ – il n'y a pas eu davantage de *psychologie* des vertus qu'il ne fallait pour résoudre des ambiguïtés importantes et répondre à des critiques. Les détails « techniques » ne sont données qu'au compte-goutte, sur des questions précises perçues comme étant problématiques, souvent isolées et en dehors d'une théorie psychologique positive et systématique

Pour la sensibilité du théoricien causal, cette façon de procéder est exaspérante et paradoxale : au lieu d'expliquer la structure psychologique (par exemple) du courage et la façon dont il produit ses bénéfices, une grande masse d'articles posent des questions spécifiques : « peut-on être courageux et juste en même temps ? », « l'individu courageux arriverait-il à avoir une réaction non-détachée avec ses proches ? », « le courage du guerrier et le courage du citoyen sont-ils la même attitude ? », etc. Ce serait comme débattre sur la possibilité qu'un vaisseau spatial qui sera développé en 2030 puisse atterrir ou non dans un désert de Mars, mais sans avoir fixé d'abord ses caractéristiques générales et son plan d'ensemble. La littérature abonde d'exemples d'absence d'explication profonde. Ce passage de Mohanan résume très bien la situation :

« Ceux qui essaient de donner quelques indications sur ce qu'est une vertu normalement finissent par formuler l'une des choses suivantes : quelques indications sur ce qu'une vertu fait mais non sur la façon dont elle fait ce qu'elle fait ; un examen des domaines de la vie morale et sociale dans lesquels les vertus opèrent ; une déclaration d'aucune utilité selon laquelle la vertu est un trait de caractère, une déclaration selon laquelle la vertu est liée à l'eudémonie, ou une ébauche d'une théorie psychologique des vertus qui ne nous en dit pas

⁴⁵ G. E. M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33, no. 124 (1958): 1.

⁴⁶ N. J. H. Dent, *The Moral Psychology of the Virtues* (Cambridge University Press, 1984), Monahan, "Meeting Anscombe's Demand: Toward a Moral Psychology of Character".

*suffisamment pour construire une psychologie morale réaliste qui serve de fondement à l'éthique de la vertu*⁴⁷. »

Philippa Foot opère une distinction entre les vertus qui aident à résister à la tentation et celles qui permettent de stimuler une motivation insuffisante. Dans le compte-rendu de von Wright, les vertus permettent de contrer l'effet négatif de la passion sur les jugements à propos de nuisance et bénéfice. McDowell définit la vertu comme la capacité à reconnaître les contraintes que la situation impose sur l'action et McIntyre définit la vertu comme la qualité de l'esprit et du caractère permettant de reconnaître et acquérir certains biens humains⁴⁸. Dans tous ces cas, les liens traditionnels entre vertu, volonté, décision, raison et connaissance sont affirmés sans une explication psychologique substantielle. Beaucoup d'attention est consacrée à la classification des vertus par rapport à des domaines différents de la vie. Le courage et l'honnêteté relèvent de problèmes différents de la vie, la gestion du danger et la répartition des ressources dans la vie relationnelle : il faut donc postuler deux vertus différentes⁴⁹. Malgré l'éclaircissement, on est laissé dans l'ignorance sur la façon dont chaque vertu obtiendrait ses résultats dans le domaine de compétence.

Il apparaît que la théorie de la vertu contemporaine est extraordinairement avare de détails sur le modèle de psychologie morale qu'elle propose : en ce sens, la théorie de la vertu ancienne, en particulier Aristote, était beaucoup plus informative d'un point de vue psychologique. Le choix de plusieurs théoriciens de la vertu néo-aristotéliens contemporains de ne suivre Aristote qu'en partie et comme simple modèle d'inspiration explique la perte de précision et de détail psychologique : dans la mesure où ils s'éloignent du modèle, il ne le remplacent pas par du contenu empirique aussi informatif. La psychologie des vertus d'Aristote est souvent présentée comme un modèle inégalé ; cela devrait être davantage une raison de honte pour les contemporains qu'une exaltation d'Aristote. Mais à quelque chose malheur est bon, car tout ce débat portant sur l'exhortation non écoutée de Anscombe se situe au niveau de la psychologie d'un agent vertueux idéal. Les participants au débat – aussi bien ceux qui se contentent d'un quiétisme wittgensteinien sur le fonctionnement de la vertu que ceux qui voudraient davantage de détails – sont tous d'accord sur la nature des précisions psychologiques attendues : elles auraient très peu à voir avec la psychologie empirique. Il s'agit de

⁴⁷ Monahan, "Meeting Anscombe's Demand: Toward a Moral Psychology of Character" 13-14.

⁴⁸ P. Carus et A. C. MacIntyre, *Dependent Rational Animals: Why Human Beings Need the Virtues* (Open Court Publishing, 2001), P. Foot, *Virtues and vices and other essays in moral philosophy* (Clarendon Press), J. McDowell, "Virtue and Reason," *The Monist* 62, no. 3 (1979), G. H. von Wright, *The varieties of goodness* (Kegan Paul, 1972).

⁴⁹ Nussbaum, "Non-Relative Virtues: An Aristotelian Approach."

dériver une psychologie morale à partir d'intuitions et contraintes normatives définissant un idéal de moralité et rationalité. Tout le monde est d'accord : la psychologie empirique n'est pas concernée. Mais alors, du point de vue de l'adéquation empirique, cette couche de détails introduits sans prendre minimalement en compte la psychologie des gens réels ne fait qu'empirer la situation. C'est peut-être une chance que le conseil de Anscombe n'ait pas été trop suivi. Mais pourquoi ne pourrait-on bâtir une psychologie normative uniquement sur la base d'intuitions portant sur la rationalité et la morale? Attribuer un rôle à la psychologie empirique dans un modèle normatif ne serait-il pas commettre une faute grave de « psychologisme »? Le prochain paragraphe est consacré à cette question.

Psychologie normative et données empiriques

Imaginez que vous possédiez une vieille maison et un terrain à la campagne. Des travaux de rénovation sont envisageables, avec deux objectifs possibles : en faire une ferme ou un hôtel. Vous préférez l'hypothèse de la ferme, mais après une longue analyse des coûts, bénéfices et opportunités, il semble clair que l'hôtel est la meilleure option. La vieille maison deviendra donc un hôtel. Cette façon de procéder est parfaitement rationnelle, pourvu que la différence au niveau des coûts/bénéfices soit plus importante que la différence de désidérabilité entre les deux options.

L'exemple semble relever d'une application trop froide de la rationalité instrumentale. L'idée de la ferme était peut-être associée à une valeur affective très difficilement négociable : c'était le projet de nos parents. C'est possible, mais si le problème est posé uniquement par le caractère « très difficilement négociable » de la valeur, l'objection ne prouve pas ce qu'elle voudrait. Un calcul coûts/bénéfices *peut* prescrire l'abandon d'une option difficilement négociable : il suffit d'avoir une alternative encore plus avantageuse. Pour sortir du cadre de la rationalité instrumentale il faudrait que la décision soit prise indépendamment des coût/bénéfices économiques. Par exemple, la construction de la ferme pourrait correspondre à la volonté de Dieu, même si cela signifie s'endetter à hauteur de plusieurs centaines de milliers d'euros.

Par rapport à des idéaux de vertu et maturation du caractère humain, il semble qu'on se trouve dans le cas de figure de valeurs radicalement non négociables. En échange de quoi pourrait-on négocier le fait d'être honnête, sage, juste, bref, vertueux ? Ce n'est pas négociable, bien sûr. S'il fallait des efforts énormes pour atteindre la sagesse et la justice après des décennies de travail sur soi-même,

cela ne rendrait l'engagement que plus urgent. Un projet normatif de construction du caractère basé sur ces valeurs, et qui ignore⁵⁰ la difficulté de l'implémentation, semble parfaitement défendable. Cependant, ce jugement est trop hâtif. Il faut réfléchir sur ce qui est impliqué dans l'acceptation d'une explication de la vertu. Dans l'absolu, il semble absurde de vouloir négocier sur l'honnêteté ou la sagesse, mais tout idéal philosophique de vertu inclut aussi un volet psychologique plutôt « technique ». Certes, une conception du courage peut être vidée de tout contenu empirique informatif. Il suffit de définir le courage comme l'attitude (quelle qu'elle soit) qui permet d'agir correctement dans des situations dangereuses. Dans cette définition, l'importance non négociable de la vertu devient tautologique. Or, la tâche de toute théorie de la vertu est celle d'avancer une explication de la façon dont la vertu fonctionne qui soit au moins minimalement informative d'un point de vue psychologique. Toute théorie de la vertu (y compris les versions qui s'exposent le moins) fait des affirmations contestables en droit sur la structure psychologique du courage, de l'honnêteté, de la justice, de la sagesse, etc. Si on laisse de côté les définitions empiriquement vides et on se réfère à des dynamiques psychologiques possibles, les affirmations traditionnelles sur la vertu deviennent facilement contestables. Par exemple, dans le chapitre sur la psychologie de l'expertise, on montre que la théorie de la vertu traditionnelle – qui conçoit la vertu comme une forme d'expertise morale – a adopté une distinction de degré et non de principe entre compétence et excellence. Quelle est la différence entre un niveau professionnel et un niveau d'excellence dans l'exercice d'une activité ? D'après les anciens, les mêmes phénomènes psychologiques qui expliquent la compétence d'un bon navigateur ou d'un bon général sont poussés à l'extrême dans le cas d'un navigateur ou général excellents (parmi les meilleurs au monde). Il y a une continuité psychologique entre les deux cas. Par contre, la psychologie contemporaine de l'expertise montre l'existence d'un clivage substantiel entre la simple compétence et l'excellence⁵¹. Pour atteindre un niveau d'excellence il faut un entraînement plus long, plus intensif et de plus grande qualité (*deliberate practice*). Ce processus produit des changements majeurs dans le système cognitif : notamment, la création de véritables « modules » consacrés à la tâche. Sans cette forme d'entraînement de haut niveau, la performance stagne sur un plateau. Etant donné la difficulté de l'idéal de la vertu, sa poursuite devrait s'inspirer davantage de la psychologie de l'excellence que de la psychologie de la compétence ordinaire. Celui-ci est un exemple de critique de choix

⁵⁰ Ignorer uniquement au sens technique de « ne pas prendre en considération pour justifier » l'idéal de vertu proposé : aucun théoricien n'ignore complètement la question de l'implémentation.

⁵¹ K. A. Ericsson, "Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance," *International Journal of Sport Psychology* 38 (2007).

« techniques » de la tradition qui n'a rien à voir avec un renoncement par rapport aux ambitions initiales ; bien au contraire.

Même s'il y a des intuitions normatives et des arguments intéressants pour attribuer des propriétés psychologiques déterminées à des vertus spécifiques (ou à l'interaction entre plusieurs vertus), il n'est pas absurde d'imaginer que la structure d'un caractère moral idéal soit en effet différente.

Cet aspect « technique » de la structure normative du caractère ne doit pas être immunisé contre une évaluation des coûts/bénéfices/opportunités dans l'amélioration de la condition humaine. Il y a bien de raisons normatives pour développer une forme de courage qui puisse interagir avec l'honnêteté de la façon spécifiée par les théories traditionnelles : ce courage serait associé à des avantages appréciables. Mais on n'est pas forcé d'accepter cette approche : si développer un courage de cette forme était très coûteux et difficile – et qu'il y avait aussi d'autres structures de personnalité plus faciles à développer et capables d'assurer une conduite adéquate – on pourrait très bien accepter de sacrifier l'idéal, sans que cela soit une « hérésie » équivalente à affirmer que, dans l'absolu, on peut négocier sur la vertu.

Dans « psychologie de la vertu » c'est la partie « psychologique » qui est négociable. Sauf si, comme on l'a anticipé, on souscrit à une méta-éthique explicitement conçue pour délivrer le résultat d'une non-négociabilité absolue des structures caractérielles vertueuses. Par exemple, si, comme Zagzebski le pense, la fondation ultime de la vertu est la ressemblance à l'être personnel suprême, Dieu, il est clair que tout détail du fonctionnement psychologique vertueux devient essentiel. Car, selon cette théorie, tout détail est décisif pour maximiser la ressemblance à la structure personnelle de Dieu⁵². Un mode d'interaction spécifique entre vertus – qui peut paraître quasiment technique dans un contexte neutre – devient par principe un point de rapprochement décisif au fonctionnement de la personne divine. Cependant, l'attitude qui prévaut aujourd'hui en éthique est celle de présenter des résultats théoriques les plus généraux possibles, dont la validité ne dépendrait pas de prémisses non partageables. Il n'est pas surprenant que dans un contexte théorique théiste on puisse aboutir aux conclusions citées : la vraie question est leur généralisabilité dans un contexte plus neutre. Même les théoriciens de la vertu théistes cherchent à élaborer une psychologie de la vertu générale, la fondation théiste n'étant qu'une option non contraignante⁵³. Les exceptions dans des contextes théoriques très *sui generis* ne constituent pas une objection décisive à ma thèse.

⁵² L. T. Zagzebski, *Divine Motivation Theory* (Cambridge University Press, 2004).

⁵³ Par exemple, voir Robert Merrihew Adams, *A theory of virtue : excellence in being for the good* (Oxford ; New York: Clarendon ; Oxford University Press, 2006).

La logique de l'objection théiste, qui rapproche vertu et fonctionnement de la personne divine, peut être reformulée dans un contexte neutre. En fait, la vertu semble impliquer un certain mode de fonctionnement psychologique, par définition. Avoir agi de façon vertueuse ne se réduit jamais à avoir fait le bon choix entre plusieurs alternatives : la façon dont on agit est essentielle pour une appréciation de la vertu. Une personne généreuse ne se limite pas à donner : son acte est associé à une attitude émotionnelle et à un ensemble de réactions psychologiques adéquates. Nous apprécions que quelqu'un réagisse *d'une certaine façon* : en tant que composant essentiel de la vertu, ce volet psychologique paraît aussi peu négociable que l'idéal abstrait de sagesse, honnêteté ou courage.

La prémisse de cette objection est correcte (la vertu est essentiellement associée à un mode de fonctionnement psychologique) mais la conclusion (une psychologie de la vertu *normative* est non-négociable) ne s'en suit pas. Comme d'habitude, le diable est dans les détails. Encore une fois, il faut voir ce que les théoriciens de la vertu mettent sur la table concrètement. Il existe un certain nombre de réactions psychologiques très centrales et « presque définitives » d'une vertu donnée : par exemple, quelqu'un qui donne avec une attitude outragée et méprisante ne peut en aucun cas être considéré comme généreux. Ne pas réagir de façon méprisante n'est pas un aspect technique négociable d'une théorie normative de la vertu. Cependant, toute théorie de la vertu, même les versions qui s'exposent le moins d'un point de vue empirique, cherche pour ainsi dire à « vendre » beaucoup plus que ce noyau essentiel. La littérature abonde en discussions très informatives sur les modes d'interaction entre plusieurs vertus ; il y a un débat sur l'unité de la vertu (la thèse selon laquelle, pour avoir une vertu, il faut les posséder toutes) ; toutes sortes d'analyses cherchent à comprendre quelle vertu est décisive pour quel rôle social (médecin, soldat, parent etc.) ou phase de la vie humaine⁵⁴. À bien y réfléchir, le but d'une théorie normative du caractère est celui d'aller au-delà de platitudes vagues sur ce qu'on apprécie dans la psychologie d'un acteur moral. C'est vrai que le choix des mécanismes psychologiques sur lesquels la théorie s'expose le plus (donc, en dehors des platitudes) est lui-même motivé par des instances normatives : les théoriciens pensent que seule leur psychologie morale de référence peut permettre la pleine réalisation de l'idéal de vertu. Mais il s'agit d'affirmations critiquables et négociables sur la base des possibilités et des ressources de la psychologie humaine réelle. Le projet d'une théorie normative du caractère qui ne prenne pas en compte les données de la psychologie empirique n'est pas tenable.

⁵⁴ J. M. Cooper, "The Unity of Virtue," *Social Philosophy and Policy* 15 (1998), J. Oakley et D. Cocking, *Virtue Ethics and Professional Roles* (Cambridge University Press, 2001), Jennifer Welchman, "Virtue Ethics and Human Development: A Pragmatic Approach," in *Virtue Ethics, Old And New*, édition S. M. Gardiner (Ithaca: Cornell University Press, 2005).

Il y a aussi une raison plus directe et plus simple à comprendre pour souligner l'importance des données empiriques. Contrairement à ce qu'on pourrait croire, tout projet normatif de psychologie morale a aussi un volet descriptif. Comment cela est-il possible, étant donné qu'il s'agit d'articuler des idéaux ambitieux ? La réponse est simple : une théorie normative de l'excellence morale doit être aussi (au moins en approximation) une théorie descriptive de l'excellence morale *réelle*, dont on a des témoignages historiques et contemporains. Même si la vertu est rare et n'a peut-être été jamais entièrement atteinte, les meilleurs modèles de conduite morale s'en sont rapprochés *plus que les autres*. Ce rapprochement doit laisser des signes potentiellement vérifiables : si une psychologie morale normative est correcte, la psychologie des modèles de conduite morale qu'on connaît devrait y ressembler.

L'idée contraire n'est pas conceptuellement incohérente : il est envisageable de justifier un idéal de transformation qui n'a jamais été ni atteint ni approché. Cependant, dans le cas de la psychologie morale, cette possibilité est extrêmement peu plausible. Le problème est épistémologique. À l'exception d'un cadre méta-éthique qui postule un accès totalement indépendant (par exemple par le biais d'une révélation, dans un cadre théiste) à un idéal de vertu morale transcendante et déconnecté de l'expérience humaine réelle, une conception idéale ne peut être plausiblement justifiée que par le fait de rendre compte de l'excellence morale réelle. Pour que le potentiel de développement moral soit intelligible, il faut en voir les signes quelque part : par exemple, on comprend la possibilité d'une sensibilité supérieure à partir de l'exemple des personnes les plus sensibles qu'on a rencontrées ; la possibilité d'une intelligence supérieure à partir de l'exemple des personnes à l'esprit plus vif qu'on a jamais connues, etc. De la même façon, un idéal de vie morale supérieure ne peut qu'être inspiré par les modèles moraux auxquels on a accès. Autrement, on serait confronté au problème de justifier le choix arbitraire de l'idéal proposé.

Il paraît évident que la psychologie philosophique de la vertu est soumise à des contraintes empiriques plutôt strictes : dans la mesure où il y a des données empiriques fiables sur les dynamiques psychologique des modèles moraux, la théorie peut être presque directement testée. Certes, le théoricien de la vertu a une certaine marge de manœuvre, essentiellement liée au fait que le modèle proposé est souvent idéalisé et ne peut être approché que par approximation et avec plusieurs limites. Mais cette marge de manœuvre est beaucoup plus limitée qu'on ne le croirait. L'existence de différents niveaux de rapprochement d'un certain modèle de psychologie vertueuse peut à son tour être quantifiée et testé expérimentalement. Ce n'est pas forcément un phénomène vague. Par exemple, dans le chapitre consacré à l'étude de la cohérence dans le comportement, on analyse des expériences dans lesquelles les psychologues se posaient la question suivante : parmi les participants, y a-t-il des sous-groupes qui tendent à manifester un comportement plus cohérent

(cohérence inter-situationnelle : la manifestation de la même attitude dans plusieurs situations pertinentes) que les autres ? Il y a plusieurs raisons possibles pour justifier la différence : par exemple, les traits qui sont plus centraux dans la conception que les individus se font d'eux-mêmes pourraient être associés à davantage de cohérence comportementale⁵⁵. L'hypothèse testée ici est l'existence d'une *tendance* à être plus cohérent que les autres, non la manifestation d'un comportement parfaitement cohérent. On voit bien que le niveau de rapprochement d'un idéal de fonctionnement psychologique peut faire l'objet de tests empiriques.

Finalement, après avoir analysé la structure théorique d'un projet de psychologie normative des vertus, il paraît que rien ne le met à l'abri de la confrontation avec des données empiriques peu favorables. La prise en compte de nos connaissances sur la structure de la personnalité telle qu'elle est étudiée par la psychologie empirique est nécessaire.

Il y a une dernière position implicite qu'on pourrait charitablement attribuer à un théoricien de la vertu qui ne se mesure pas de façon approfondie aux données empiriques. Il est possible qu'il accepte entièrement la vulnérabilité de son projet par rapport aux découvertes empiriques ; néanmoins, il juge implicitement qu'aucune découverte empirique sur le fonctionnement de l'esprit humain ne menace le projet philosophique. Dans cette hypothèse, aucune révision ne serait nécessaire. Il n'y aurait pas d'autres idéaux à préférer du fait des limites et des opportunités propres à la nature de la psychologie humaine.

Si le désaccord n'est pas méthodologique et réside uniquement dans une évaluation différente (implicite, dans le cas du philosophe) d'un très grand corpus de littérature psychologique, la réponse ne peut que se placer sur le même terrain. Le but de plusieurs chapitres de cette thèse est celui de montrer concrètement que les découvertes de la psychologie sociale et de la personnalité posent plusieurs problèmes à une théorie de la vertu classique conçue comme une théorie des traits.

⁵⁵ Voir le chapitre « A la recherche de la cohérence du comportement », paragraphe « Méta-traites ».

Pourquoi et comment étudier le sens commun ?

Introduction et résumé - Le sens commun, ce grand inconnu

La référence au sens commun est une stratégie courante dans le discours quotidien et dans l'élaboration théorique, scientifique et philosophique. Pourtant, depuis quelques décennies l'intelligibilité de la notion de sens commun fait l'objet d'attaques tous azimuts. Deux problèmes se posent : pourquoi devrions-nous nous intéresser au sens commun ? À quoi se réfère la notion elle-même?⁵⁶. Toute référence au sens commun est devenue plus coûteuse à cause du premier problème et plus suspecte à cause du deuxième. Loin d'épargner les théories normatives de la personnalité, le statut problématique nouvellement acquis par « le sens commun » les touche de plein fouet. Comme on le verra dans le prochain chapitre, consacré à l'étude de la théorie du caractère du sens commun, celui de la personnalité est un domaine de réflexion dans lequel les emprunts au sens commun sont systématiques et stratégiques.

Dans ce chapitre, on cherchera à répondre au pourquoi et au comment de l'étude du sens commun. En ce qui concerne le pourquoi, on va défendre une approche plutôt traditionnelle : les prises de positions normatives du sens commun ne sont pas forcément correctes, mais elle constituent – méthodologiquement et stratégiquement – un point de départ incontournable pour une théorie normative. Tout comme un révolutionnaire cherche à changer le monde à partir des forces et ressources dont il dispose, un théoricien normatif ne saurait ignorer les engagements normatifs qui existent déjà dans le sens commun. On appelle cette position « populisme normatif ». Elle s'oppose au « fiabilisme », l'idée selon laquelle il ne faut pas partir des intuitions du sens commun, mais des intuitions des philosophes (censées être différentes).

D'après le fiabilisme, défendu par des auteurs tels que Sosa et Williamson, les intuitions des philosophes sont aux questions normatives ce que les intuitions des ingénieurs sont aux questions techniques, ou les intuitions des généraux à la stratégie militaire. La comparaison est bien choisie, car la connaissance des experts est largement intuitive, la psychologie contemporaine l'a prouvé. Toutefois, alors que les bons généraux gagnent la guerre et les bons ingénieurs construisent des ponts qui durent, il n'y a pas de preuve indépendante de la validité des intuitions normatives des

⁵⁶ La deuxième question n'est pas loin d'être paradoxale : devrait-on dire que celle de sens commun n'est plus une notion claire et familière, elle ne fait plus partie du...sens commun ? Les notions du sens commun, y compris la notion de sens commun elle-même, sont certes familières mais loin d'être claires et adaptées à un usage théorique.

philosophes. Si on prend l'exemple de la morale, pour défendre une position fiabiliste, il faudrait envisager les philosophes comme des sages ou des saints qui ont fait des choix moraux particulièrement réussis. Cela est peu crédible.

Dans la première section, on défend une approche populiste du pourquoi du sens commun : la complexité de la question « comment découvrir les positions du sens commun ? » nous occupera pour le reste du chapitre. Dans la première section, on écarte également l'hypothèse selon laquelle les philosophes seraient en mesure d'interpréter correctement la position des gens, sans l'aide de la psychologie expérimentale. Les philosophes sont en contact avec la réflexion des gens, mais cela n'est pas suffisant. Si les philosophes réfléchissent longuement sur une question, jusqu'à stabiliser des intuitions dites « réfléchies », ils sont probablement en train d'altérer la position originale des gens, de la même façon que n'importe quelle sous-culture produit une vision différente et originale de la réalité.

S'il s'agit d'interpréter de façon non superficielle la position du sens commun, mais sans l'altérer – pour arriver aux « intuitions robustes » –, les philosophes ont sans doute des compétences intéressantes. Ils peuvent interagir et dialoguer de façon profonde avec les gens. Mais ce processus peut être influencé par toutes sortes de biais, et, en termes d'efficacité, n'est pas comparable à la panoplie de moyens expérimentaux dont la psychologie dispose pour étudier l'usage des concepts ordinaires. Ces moyens sont loin d'être tous quantitatifs et « froids », comme le philosophe le croit (Kauppinen). La psychologie a intégré depuis des décennies des méthodes qualitatives et « chaudes », telles que des interviews dans lesquelles des évaluateurs expriment leurs impressions selon des codes et des critères objectivement vérifiables et intersubjectivement comparables.

Une fois compris que les compétences des philosophes et des psychologues doivent être réunies pour étudier le sens commun, on affrontera les problèmes qui empêchent une compréhension adéquate des engagements théoriques du sens commun. Un populiste normatif s'intéresse aux engagements des gens, mais uniquement s'ils sont profonds et stables. Quel intérêt y aurait-il à prendre en considération, pour une théorie philosophique, le contenu d'engagements qui pourraient changer le lendemain ? La nécessité d'étudier des engagements stables introduit une dimension holiste et diachronique dans l'étude du sens commun. Des intuitions stables et profondes sont enracinées dans l'« écologie cognitive » des gens : pour tester stabilité et profondeur il faut forcément avoir une vision d'ensemble du fonctionnement de la cognition ordinaire. Par exemple, il n'est pas possible d'étudier la théorie morale du sens commun sans une bonne compréhension du fonctionnement de la cognition sociale dans son ensemble. On ne peut pas se contenter d'enregistrer une liste d'intuitions, aussi stables et enracinées qu'elles *semblent*, ou d'étudier des aspects locaux du fonctionnement de la cognition ordinaire.

Ce manque de holisme intéresse aussi bien les philosophes que les psychologues. Les premiers ont tendance à se concentrer sur une liste de jugements intuitifs et à négliger l'étude de la manière dont les gens raisonnent. Les deuxièmes prennent en considération simultanément les jugements intuitifs et la façon dont les gens raisonnent, mais sur une échelle souvent trop locale.

Chacun étant concentré sur une liste limitée de tâches et de protocoles expérimentaux, ils défendent et projettent sur la totalité de la cognition sociale les résultats de ce qui émerge dans leur mini-domaine (formation des impressions, schémas, stéréotypes, attribution de croyances et désirs, etc.). Une étude précise et rigoureuse de questions très bien délimitées est essentielle pour le développement de la psychologie empirique, il n'y a aucune critique là-dessus. En plus, les psychologues ne soutiendraient jamais explicitement que les stratégies de fond de la cognition sociale peuvent être étudiées à partir d'un ensemble limité de tâches. Pourtant, cette fragmentation est bien l'effet émergent. C'est comme si chaque syndicat défendait le rôle essentiel dans la société de la profession qu'il représente, sans qu'une vraie hiérarchie des priorités puisse émerger. Il faut mettre en perspective le rôle de croyances, désirs, intentions, traits, stéréotypes sociaux, de processus de formation d'impression ou de la consolidation des connaissances que nous avons des êtres les plus proches. On a besoin d'une véritable cartographie quantitative et qualitative du fonctionnement des différentes ressources de la cognition sociale.

Une deuxième difficulté est posée par l'imbrication entre la fonction descriptive et la fonction pragmatico-stratégico-manipulative de la cognition sociale. Dans les interactions sociales, on assiste tout le temps à une négociation du pouvoir, de rôles et de réputation. Toute affirmation, attribution, inférence ou explication nous fait avancer ou perdre des points dans ce jeu. La cognition sociale est programmée pour nous permettre en même temps de déchiffrer de façon correcte la réalité sociale et de jouer cette partie dans les dynamiques complexes du groupe. Pour comprendre quels sont les véritables engagements théoriques du sens commun, il faudra passer au peigne fin les dynamiques de la cognition sociale, pour éliminer ces interférences pragmatiques. Car plusieurs aspects de l'usage quantitatif et qualitatif des concepts (ex : est-ce qu'on mobilise le concept de trait plus que le concept de croyance ? dans quel contexte l'un est préféré à l'autre ?) ordinaires répondent à des exigences pragmatiques et n'ont rien à voir avec la « théorie de l'esprit » cautionnée par le sens commun. Par exemple, si ce n'était pour des raisons pragmatiques, certains concepts seraient mobilisés beaucoup plus (ou moins) et d'en d'autres contextes. Cela montrerait leur véritable importance théorique (ou manque d'importance), dans le cadre de la théorie implicitement cautionnée par le système cognitif.

Pour résumer, l'étude du sens commun doit être justifiée de façon populiste (il est stratégiquement et méthodologiquement important de partir des engagements profonds des gens) et conduite de

façon étroitement interdisciplinaire entre philosophie et psychologie expérimentale. En ce qui concerne la cognition sociale, il faut dépasser la fragmentation en mini-domaines pour arriver à une vision d'ensemble, qui doit être soigneusement interprétée et évaluée pour éliminer les interférences pragmatiques et dégager les véritables engagements ontologiques du sens commun.

Section I – l'étude du sens commun en philosophie – l'insuffisance des approches traditionnelles

Quelle est la position du sens commun ? L'importance de la philosophie expérimentale

Avant d'aborder le sujet spécifique de la théorie du caractère du sens commun dans le prochain chapitre, il est essentiel d'avoir assimilé la leçon suivante : découvrir les positions implicites (et même explicites, d'ailleurs) du sens commun sur des questions importantes d'un point de vue philosophique est une entreprise très difficile et largement empirique. Soulever directement le problème de la position du sens commun sur le caractère, en dehors de ce débat plus large, ne saurait être ni convaincant ni satisfaisant. Dans cette section, on cherchera à défendre d'un point de vue méthodologique la nécessité d'études empiriques très approfondies contre l'approche de la « philosophie du canapé » (*armchair philosophy*). Etant donné notre centre d'intérêt – la vertu – on se concentrera sur la position du sens commun sur des questions de philosophie morale, mais les conclusions sont également valables dans tous les autres domaines.

S'il y a dix ans, on avait dit que la philosophie morale (et en général la philosophie normative : théorie de la rationalité ou de la connaissance) se basait sur les intuitions du sens commun, on aurait sans doute affirmé quelque chose d'évident. Nul n'aurait contesté cette affirmation. Une série d'*intuitions* et de « prises de positions » plus générales du sens commun étaient considérées sans doute comme un outil de travail du philosophe normatif. « Ne pas tuer une personne pour en sauver trois », « on ne sait pas qu'on ne gagnera pas à la loterie », « faire activement du mal est pire que ne pas intervenir et laisser le mal se produire » sont autant d'exemples de cette matière première du travail du philosophe.

Mais qu'est-ce qu'un jugement intuitif ? Dans la tradition qui a dominé la philosophie de langue anglaise du vingtième siècle, une intuition est, à un moment donné, un jugement immédiat qu'une personne formule sans avoir l'impression de réfléchir ou de connaître une justification plus

articulée. Cette justification plus profonde peut être recherchée. Toutefois, si le jugement est intuitif d'un point de vue psychologique, alors, au moment concerné, on ne le formule pas *en vertu* de la croyance en cette justification. Par exemple, on peut formuler une théorie de la connaissance qui est incompatible avec la possibilité de savoir qu'on n'a pas gagné à la loterie. Mais les gens n'ont pas cette intuition (il est impossible de savoir qu'on n'a pas gagné la loterie) *parce qu'ils croient* dans cette théorie.

Cela dit, il a été toujours clair que tous les éléments qui entrent dans la galaxie «intuitions du sens commun» n'ont pas le même statut : il y a sans doute au moins une différence énorme entre des jugements intuitifs de premier niveau sur des cas d'application déterminée d'un concept (il n'est pas «juste» de tuer une personne pour en sauver trois, dans le cas *x* déterminé) et des principes généraux intuitifs tels que «tuer est pire que laisser mourir» (d'ailleurs, tous les principes de ce genre n'ont pas forcément le même niveau de généralité). Beaucoup de distinctions et classifications intéressantes sont possibles, mais, il y a dix ans, la question se serait plutôt vite déplacée de «qu'est-ce qu'une intuition?» au problème de comment l'exploiter au niveau théorique.

Il n'y a jamais eu d'accord sur la façon d'exploiter les prises de positions du sens commun, l'éventail d'options méthodologiques s'étalant sur un continuum qui va de l'exigence que les théories philosophiques normatives les respectent le plus possible⁵⁷ à la possibilité de les rejeter en masse⁵⁸, passant par une série de positions méthodologiques qui permettent de négocier le poids des intuitions les unes contre les autres⁵⁹ ou contre d'autres sources⁶⁰.

Avant même que la philosophie expérimentale n'entre dans l'arène, un débat aussi intense sur ce qu'il faut faire du sens commun a complexifié un peu l'image de ce dernier : au-delà de la

⁵⁷ Il faut rejeter le moins possible d'intuitions, et parmi les moins importantes. On juge les théories concurrentes par leur capacité à préserver le sens commun. Les intuitions jouent ici le même rôle que les observations empiriques en physique, avec l'impératif de sauver les phénomènes.

⁵⁸ Il s'agit d'une version radicale de la théorie de l'erreur, introduite par John Mackie : pour des raisons indépendantes, par exemple métaphysiques, on considère que les propriétés morales telles que «juste» ou «bon» sont incohérentes, le sens commun est donc systématiquement en erreur en jugeant que telle chose est bonne ou juste. Voir J. L. Mackie, *Ethics: Inventing Right and Wrong* (Penguin, 1977).

⁵⁹ C'est la célèbre approche de l'équilibre réfléchi popularisée par Rawls : on cherche le meilleur équilibre entre des intuitions de premier niveau sur des cas déterminés et les principes intuitifs. Dans certaines interprétations, pour atteindre un équilibre stable il peut être nécessaire de «sacrifier» un bon nombre d'intuitions de départ

⁶⁰ Ce qui a été appelé équilibre réfléchi «large» : on a le droit de négocier le poids de nos intuitions prenant en compte aussi la volonté de préserver des théories juridiques importantes, ou d'avoir une théorie finale «harmonisée» avec les sciences empiriques. Voir J. Rawls, *Théorie de la justice* (Paris Le Seuil, 1987).

distinction entre principes généraux intuitifs (plus ou moins généraux) et intuitions, parmi ces dernières, on peut distinguer entre intuitions *prima facie* ou immédiates et intuitions réfléchies (des jugements qui restent et « se stabilisent » après qu'on a réfléchi longuement sur eux). Cela donne déjà un nombre intéressant de combinaisons : on peut être un ennemi ou un défenseur acharné du sens commun dans ses intuitions de base, dans ses principes ou dans ses intuitions réfléchies, ou à plusieurs de ces niveaux.

Cependant, ces distinctions ne représentent qu'une minorité de moments particulièrement lucides et auto-réflexifs dans la pratique philosophique du vingtième siècle : la plupart des théoriciens ont avancé leurs travaux en présentant des prises de positions comme relevant de « ce que pense l'homme ordinaire », « de l'opinion répandue », bref du sens commun, sans qu'on puisse suivre toujours au fil de l'argumentation de quel genre d'entité il était question. Finalement, aussi problématique soit-il de décider qu'en faire, le sens commun a toujours été perçu comme une matière première indispensable et surtout *clairement identifiable*, tout comme les maçons s'accordent sur ce qui est un marteau et une brique, quoi qu'ils veuillent construire.

L'attaque impromptue et récente de la philosophie expérimentale, qui accuse les philosophes d'être incapables de décrire de façon adéquate le sens commun sur lequel ils s'appuient, a changé complètement la donne. On sait que les suppositions paradigmatiques de fond d'une approche théorique ne deviennent claires que lorsque le paradigme est déjà en crise, et que jamais sa propre position ne devient aussi précise que lorsqu'on est obligé de la reformuler sous la pression des critiques. Ce phénomène s'est produit de façon remarquable dans le cas du rapport traditionnel entre philosophie et sens commun, qu'on commence à comprendre avec un niveau acceptable de clarté seulement lorsqu'il est peut-être temps de le reformer définitivement. Ainsi, on ne peut commencer que par une question d'éclaircissement : qu'est-ce que réellement le sens commun pour un théoricien normatif et pourquoi s'y intéresse-t-il ?

« Populisme normatif et expertise » vs « fiabilisme et révisionnisme éclairé »

En appelant cela sens commun (qu'ils aient raison ou non de le faire), les philosophes en pratique font confiance à leurs intuitions (même si, ensuite, il s'agissait de les rejeter par une théorie de l'erreur). Du point de vue logique, on peut y objecter deux choses : (1) qu'il ne faille pas faire confiance à quelque chose comme des intuitions tout court, ou (2) qu'on ne doive pas se fier aux intuitions des philosophes. Ces derniers, pour leur défense, doivent fournir deux explications : quelle est la plausibilité initiale d'une référence à des intuitions *en général*, et pourquoi leurs intuitions et non pas celles d'un autre groupe socio-économique ou culturel ?

Dans la littérature stimulée par la critique des philosophes expérimentaux, deux réponses à la première question ont émergé, chacune assortie d'une réponse spécifique à la deuxième. La plupart des philosophes ont défendu une version de « populisme normatif » (1^{ère} question) associé à une doctrine de « l'expertise » des philosophes dans la tâche d'identifier les vraies intuitions des gens (2^{ème} question). En un mot, le populisme normatif est l'idée selon laquelle, pour bâtir une théorie normative, il est stratégiquement avisé de partir des engagements les plus profonds et basiques que les gens ont *déjà*. Cette approche n'est pas uniquement majoritaire, elle est aussi fidèle à l'histoire des disciplines normatives du vingtième siècle, car elle représente vraisemblablement l'attitude implicite dans le débat réel.

La deuxième approche à la première question (pourquoi faire confiance à des intuitions, tout court ?) est le « fiabilisme » : si le thème du populisme normatif est le caractère incontournable de nos engagements profonds manifestés par les intuitions, quitte éventuellement à les rejeter comme incohérents à la fin, le fiabilisme défend l'idée selon laquelle il faut faire confiance aux intuitions tout simplement parce qu'elles sont fiables. Au moins une partie des intuitions sont des jugements fiables, lesquelles ? Ici le fiabilisme est assorti du « révisionnisme éclairé » (2^{ème} question : pourquoi les intuitions des philosophes ?) : il s'agit de l'idée selon laquelle les philosophes ne sont pas des interprètes extraordinaires de la pensée profonde des gens, mais, au contraire, des réformateurs conceptuels, qui, en vertu de leur autorité, diffusent leur intuitions et proposent des modifications dans les concepts communs. Pour des raisons à préciser, il faudrait ainsi « faire confiance » aux intuitions des philosophes et modifier les siennes, car les premières sont plus fiables. Cette approche n'est pas uniquement révisionniste dans sa substance, mais aussi par rapport à la pratique traditionnelle et courante des philosophes.

Au niveau stratégique, d'après le populisme normatif, la plausibilité initiale d'une référence au sens commun dans le cadre d'une théorie normative se justifie par le constat suivant : de quoi faudrait-il partir pour savoir ce qu'il faudrait faire ou penser, sinon des engagements plus profonds et basiques que les hommes *ont déjà* ? Aucun philosophe n'a pensé que mettre en évidence le niveau plus basique et élémentaire de nos engagements factuels pouvait résoudre automatiquement le problème de la normativité (après tout, même la méthode la plus conservatrice admet qu'il faut rejeter certaines intuitions du sens commun), mais d'où partir si ce n'est par ce qui nous engage déjà profondément ?

Si, en morale, on ne peut pas partir du constat que (par exemple) « il est absurde de tuer une personne innocente pour en sauver trois déjà condamnées » ou d'autres jugements qui nous paraissent également évidents, ne se destine-t-on pas à un vide et à un arbitraire méthodologiques

absolus ? Insister que la morale est arbitraire et que la méthodologie pour l'étudier ne peut (par conséquent) que l'être aussi n'est pas une objection pertinente, car l'argument du philosophe « populiste » concerne les intérêts stratégiques et méthodologiques de quelqu'un qui *veut* essayer de développer une théorie normative solide : que la seule approche sensée commence par prendre en considération le sens commun ne préjuge pas du succès final de l'entreprise. Il peut être vrai qu'on a au mieux un seul chemin à essayer, avant de découvrir qu'il n'y a pas de solution. Pour cette raison, un pessimisme ou un scepticisme de base quant au projet d'une théorie normative ne touchent pas à la crédibilité du projet populiste.

Ainsi, il s'agit moins de critiquer la motivation stratégique du populiste que de voir plus clair dans la réalisation de ce projet. Pour qu'il puisse s'agir de quelque chose d'utilisable dans cette logique, par « sens commun » on doit se référer à des jugements intuitifs qui ont ces trois caractéristiques :

ils doivent être universels (ou *presque*, voir ci-dessous pourquoi) ;

ils doivent être facilement accessibles et reconnaissables par un philosophe (au moins comparativement à d'autres jugements ou doctrines qui ne relèvent pas du sens commun) ;

il est difficile de les réviser ou rejeter. Ils représentent des engagements profonds.

D'abord, une série de clarifications sur ces contraintes. Prenons un cas typique : pour que l'intuition « toutes choses étant égales par ailleurs, il est injuste de tuer un individu innocent et pas concerné, pour sauver deux individus condamnés » soit un jugement du sens commun dans le sens pertinent, il faut que presque tout le monde le pense et que ce consensus soit plus ou moins évident et facile à établir par le philosophe. Il faut aussi que nous soyons profondément engagés par rapport à ce jugement. La troisième contrainte paraît triviale car, bien sûr, si le théoricien populiste s'intéresse au jugement intuitif « tuer est pire que laisser mourir » c'est que le sens commun ne change pas d'idée tous les jours sur cela. Mais elle ne l'est pas. D'abord, par principe : étudier le sens commun n'est pas faire un instantané d'un ensemble d'attitudes, c'est depuis le début une question *diachronique* et *contrefactuelle*. Le problème du sens commun intègre la question suivante : comment le sens commun évoluera ou pourrait évoluer face à certains inputs ? Ensuite, pour des raisons méthodologiques : une chose est d'être sûr de la manière dont les gens raisonnent, autre chose d'être sûr de la manière dont les gens vont raisonner ou raisonneraient face à certaines sollicitations (que vérifier les deux choses soit difficile ou non). Il faut faire référence à des remarques ou à des observations différentes.

En plus, le caractère diachronique et contrefactuel va de pair avec une référence à des processus cognitifs profonds et universels. Souvent tout se passe comme si les philosophes constataient l'existence de jugements communs « de surface » qui constituent leur point de départ, exactement dans la similitude avec les observations en physique : « on ne sait pas qu'on ne gagnera pas à la

loterie », « on ne sait pas qu'on n'est pas un cerveau dans une cuve », et celles-ci sont les contraintes pour bâtir une théorie épistémologique, tout comme la physique doit sauver les phénomènes. Mais la comparaison ne tient pas, car si le philosophe s'intéresse forcément à la façon dont les gens raisonneraient, a fortiori il s'intéresse aux processus de raisonnement, aux mécanismes cognitifs profonds qui expliquent les jugements « superficiels » dont il est question. Peu thématifiée dans l'auto-compréhension méthodologique des philosophes, cette attention au cognitif (au moins dans le sens neutre et non controversé de « comment les gens raisonnent ») a été pourtant fondamentale depuis le début dans la tradition de l'analyse conceptuelle, par rapport au problème de l'explication de l'erreur.

Nous venons de dire que les intuitions prises en considération par le philosophe doivent être universelles ou presque. Cela requiert deux clarifications : la première est que le « presque » ne se réfère pas aux dimensions de l'ensemble de référence pris en considération par le philosophe. En d'autres termes, nous n'avons pas dit « presque » parce qu'un philosophe « culturaliste » pourrait limiter son discours au sens commun de la tradition occidentale. Cette possibilité, dont la justification ultime ne nous intéresse pas dans le cadre d'une analyse de la plausibilité initiale d'une stratégie, n'affecte pas notre discussion : on peut bien relativiser l'universalité à un ensemble de référence, parler par exemple d'intuitions et stratégies cognitives d'une culture, plutôt que de l'espèce.

La deuxième clarification est la suivante : les intuitions ne doivent pas être universelles dans l'ensemble de référence, une bonne intuition n'est vraiment universelle que *potentiellement*, si la personne concernée réfléchit dans des conditions normales (ou idéales) et ne fait pas d'autres fautes ou confusions conceptuelles. Autrement dit, alors que la source ultime de plausibilité stratégique est ce sur quoi « nous » (dans un ensemble de référence) sommes déjà engagés, le sens commun peut bien se tromper localement à cause (banalement) de manque d'attention ou de confusions superficielles, ou (de manière plus intéressante) de confusions conceptuelles profondes. Ainsi, avant de recourir à la solution ultime de rejeter une intuition dont on reconnaît qu'elle appartient au sens commun, on cherche à expliquer pourquoi, à la suite de quelles erreurs ou confusions conceptuelles, certains sujets semblent ne pas l'avoir. Les exemples sont innombrables, ils vont de pourquoi les gens n'ont pas les bonnes intuitions utilitaristes (par exemple, ils prennent trop au sérieux des principes qui poussent aux bons choix utilitaristes seulement en général, sans voir qu'il faudrait faire des exceptions) à toutes les confusions conceptuelles dénoncées dans la tradition wittgensteinienne (les gens croient que les états mentaux sont dans le cerveau parce que dans certains usages linguistiques on parle d'*avoir* une croyance, comme s'il s'agissait d'une chose). La

thérapie wittgensteinienne n'est pas autre chose que permettre au sens commun de retrouver ses bonnes intuitions⁶¹.

Tout ceci, en ce qui nous concerne, montre comment depuis le début la référence au sens commun n'a pas de sens si on ne la conçoit pas comme une étude d'un répertoire conceptuel et de stratégies cognitives dynamiques universelles (dans un ensemble de référence et au niveau potentiel : ces stratégies peuvent être perturbées par des confusions conceptuelles ou d'autres problèmes). La remarque est importante, car les philosophes s'engagent dans quelque chose de beaucoup plus fort qu'il ne paraît souvent. Ils ne peuvent pas s'en tirer par l'attitude suivante : « ok, on ne sait pas qu'on ne va pas gagner la loterie, qui dirait le contraire ? On peut bâtir notre théorie à partir de ce constat », car toute référence au sens commun postule l'existence d'attitudes inférentielles universelles et permet de faire des prédictions contrefactuelles très précises sur comment les gens raisonneraient, ce qui est extrêmement ambitieux.

Dans notre exemple, l'épistémologue en question ne se limite pas à prendre note d'opinions répandues. Il s'engage à soutenir que (1) les gens continueraient à penser qu'on ne sait pas qu'on ne gagnera pas à la loterie, même face à de nouveaux inputs et objections ; (2) les gens qui ne le pensent pas le penseraient après la « thérapie » adéquate, après avoir détecté la confusion conceptuelle qui cause le problème. Ce sont des postulats beaucoup plus coûteux que cela n'en avait l'air au premier abord. Coûteux, et surtout apparemment très *empiriques*, car si étudier comment les gens raisonnent et faire des prédictions précises sur comment ils raisonneraient ne relève pas de la psychologie, on ne voit pas ce que les gens dans les départements de psychologie seraient censés faire. L'objection selon laquelle ces intuitions sont censées être normatives (elle ne décrivent pas des faits, elle portent sur le devoir être) et que la psychologie n'étudie pas des questions normatives, n'est pas pertinente, car, on l'a vu, la justification populiste, pour prendre au sérieux la portée normative des intuitions, est que tout le monde les partage *dans les faits*. Le constat descriptif précède le normatif. Et pourtant, les philosophes défenseurs de la philosophie du canapé (*armchair philosophy*) soutiennent qu'ils sont les mieux placés pour vérifier si tous ces processus ont lieu ; ils seraient même des experts en la matière.

La nécessité de la philosophie expérimentale, dans l'approche populiste

Les philosophes expérimentaux, en plus d'avancer et tester des hypothèses spécifiques dans chacun de leurs articles, d'un point de vue méthodologique ne font que soutenir la nécessité d'explorer le

⁶¹ M. Budd, *Wittgenstein's Philosophy of Psychology* (London: Routledge, 1989).

sens commun avec des expériences psychologiques de plus en plus complexes et raffinées. C'est, d'après eux, une question difficile et surtout empirique. Pour cela, il ne sera pas pertinent de discuter les objections des philosophes qui trouvent des problèmes spécifiques dans le paradigme expérimental utilisé dans une série d'expériences déterminées : accuser un psychologue de n'avoir produit qu'un artefact est le pain quotidien de la critique interne qui alimente la recherche psychologique. À la production d'artefacts, on répond avec davantage de recherches plus raffinées, non pas avec moins. Il faut s'intéresser aux raisons méthodologiques de principe que les philosophes ont avancées pour montrer l'avantage de l'approche traditionnelle sur l'idée d'une étude empirique du sens commun.

L'idée de l'expertise des philosophes dans l'analyse des concepts et de la pratique inférentielle des gens a été défendue par deux stratégies différentes, une qui se base sur la distinction déjà introduite entre intuitions immédiates et intuitions réfléchies, et l'autre qui se base sur la distinction différente entre intuitions superficielles et intuitions robustes. En général, la nécessité d'expertise s'impose parce qu'il n'est pas plausible que tout le monde ait une vision adéquate de la manière dont les gens pensent et raisonnent. En psychologie, l'influence de biais tels que le biais de confirmation,⁶² de disconfirmation⁶³ et de faux consensus⁶⁴ a été abondamment démontrée. Cela, toujours d'après les données, est loin d'épargner des groupes spécifiques de population, y compris les experts dans un certain domaine. Si la nécessité d'expertise dans l'étude des concepts commun est claire, dans le sens où poser directement des question à un échantillon de gens dans la rue ne serait pas adéquat, on ne voit pas, sans des raisons spécifiques, pourquoi on devrait faire confiance aux philosophes, d'autant plus qu'ils sont un groupe restreint avec une formation très particulière (cela dit, pour l'instant il n'y a pas non plus des expériences qui prouvent qu'ils sont des victimes de biais plus systématiques que les autres).

Les philosophes en tant qu'experts dans l'étude des intuitions réfléchies

La première stratégie consiste à dire que les seules intuitions qui comptent pour révéler les engagements profonds du sens commun sont les intuitions réfléchies, celles qui se stabilisent après

⁶² Attention sélective aux données qui viennent à soutenir de son propre point de vue, voir Mynatt *et al.* 1977; Nickerson 1998; Wason 1960.

⁶³ Attention sélective aux critiques qui vont à l'encontre de sa propre position, par exemple chez des gens qui se considèrent persécutés, voir Ditto & Lopez 1992; Edwards & Smith 1996

⁶⁴ Par exemple, on n'arrive pas à croire qu'un candidat a gagné les élections parce que personne parmi nos amis n'aurait voté pour lui. Voir Ross *et al.* 1977; Fields & Schuman 1976.

une longue période de réflexion. Cela ne leur enlève pas leur statut d'intuitions, car une intuition se définit par la phénoménologie de son expression à un moment donné : quelque chose d'intuitif n'est pas nécessairement non-affecté indirectement par mes réflexions passées. Étant donné que l'homme lambda ne réfléchit pas forcément longtemps sur ses engagements basiques, il est plausible qu'il ait très peu d'intuitions réfléchies : avoir des intuitions réfléchies semble par contre le métier du philosophe.

Cette stratégie de défense ne paraît pas adéquate, pour plusieurs raisons : d'abord, si le but est de défendre la méthodologie qui a animé la plupart des débats du vingtième siècle, il semble tout simplement faux que les intuitions mobilisées ont toujours été réfléchies plutôt qu'immédiates. Cela paraît crédible dans le contexte d'usage classique de la méthode de l'équilibre réfléchi par Rawls : l'élaboration d'une théorie de la justice. Il est plausible qu'on se réserve le droit de réfléchir pour décider si un certain modèle d'allocation des ressources est injuste. Mais, par exemple en théorie de la connaissance, souvent l'appel à l'immédiateté de l'intuition est plutôt évident, voire même explicite. C'est le cas de la discussion du scepticisme ou du problème de Gettier : qu'on ne puisse pas savoir que « nous ne sommes pas des cerveaux dans une cuve » est vendu comme quelque chose d'absolument évident. Même chose pour le fait qu'on ne puisse pas savoir que « notre voiture est garée où on l'a garée ce matin – si elle est bien là, mais qu'un voleur l'a prise et vient de la garer au même endroit il y a cinq minutes ». Cela fait même partie de la stratégie du sceptique que d'avoir des présupposés les moins coûteux possibles : dans l'interprétation traditionnelle, un défi sceptique radical doit se baser sur des constats qui ont des suppositions moins problématiques que la tentative de réponse censée le contrer. Ce serait un mauvais départ s'il s'agissait de dire que la plupart des sages qui ont évalué la question sont arrivés à la conclusion qu'on ne peut pas savoir qu'on n'est pas un cerveau dans une cuve⁶⁵.

Deuxièmement, d'après ce que nous savons du fonctionnement des biais psychologiques originellement responsables de la difficulté de l'étude des intuitions partagées, le fait de réfléchir longuement à propos de ce qu'on pense vraiment, dans un milieu isolé et très *sui generis*, en contact uniquement avec des collègues et des étudiants déjà conditionnés ou « acculturés » ne paraît pas en mesure d'aider. On a vu que le biais de faux consensus n'épargne pas les experts dans leur domaine : dès qu'on sait quelque chose, on n'arrive plus à comprendre comment les autres peuvent ne pas comprendre, et on finit par surestimer la diffusion des compétences concernées.

Si le but est d'étudier les concepts du sens commun, il semble aussi qu'une longue réflexion où on cherche à mettre à l'épreuve la stabilité d'une intuition dans son propre esprit empire la situation au

⁶⁵ D. H. Pritchard, "Recent Work on Radical Skepticism," *American Philosophical Quarterly* 39, no. 3 (2002).

lieu de l'améliorer. Un tel processus a l'air de pouvoir changer et mettre à jour son propre réseau conceptuel, alors que le but était de décrire les intuitions répandues. Telle mise à jour peut être en soi fort intéressante : même les philosophes qui idolâtrèrent le sens commun justifient le rôle de l'analyse conceptuelle par la nécessité de résoudre, par de petites modifications, les paradoxes et les contradictions présentes dans notre schéma conceptuel.

Le problème du scepticisme radical est présenté comme un cas typique de dilemme du sens commun : le rejet sceptique de la possibilité de la connaissance (nous ne savons rien) est bien sûr contre-intuitif, mais l'argument sceptique radical censé le soutenir se base sur des prémisses intuitives, notamment que (1) « on ne sait pas qu'on n'est pas un cerveau dans une cuve » ; (2) « pour une proposition p quelconque, si on sait p, alors on sait qu'on n'est pas un cerveau dans une cuve », d'où par *modus tollens*, on en dérive que « pour toute proposition p, on ne sait pas que p ». S'il y a une conclusion contre-intuitive de prémisses intuitives, c'est qu'il faut rejeter au moins une intuition, d'où la nécessité de modifier en partie le schéma conceptuel du sens commun. Le philosophe devient alors un réformateur, ce qui est tout à fait passionnant, mais incompatible avec la méta-justification du populisme normatif. Réfléchir à fond sur le fait d'avoir ou ne pas avoir une certaine intuition est la meilleure façon de *changer*, non pas d'accéder à l'intuition de départ. On verra ci-dessous que, pour exploiter le potentiel des intuitions réfléchies, il faut mobiliser une autre méta-justification du recours aux intuitions : le fiabilisme.

Les philosophes en tant qu'experts des intuitions « robustes »

Restant dans un cadre populiste, la deuxième stratégie de défense des philosophes est beaucoup plus raisonnable et en ligne avec la méthodologie de la psychologie. D'un côté les affirmations explicites du sens commun ne sont pas fiables à cause de biais, de problèmes d'attention, de confusions et interférences pragmatiques ; de l'autre côté, une invitation à une réflexion trop approfondie change la donne de départ. La solution est une étude méthodologiquement sophistiquée de la réaction des gens à des stimulations et manipulations sur une durée courte ou moyenne, de façon à ce qu'on ne s'arrête pas à la surface, sans pour autant produire une acculturation. Le psychologue des concepts est en accord complet : littéralement obsédé par la peur d'influencer le participant à l'expérience, il cherche néanmoins à dépasser la surface avec des variations et des manipulations expérimentales les plus sophistiquées possible.

Mais le philosophe pense avoir un avantage. Dans un article récent de critique de la philosophie expérimentale, Kauppinen défend la supériorité de la méthode appelé « dialogue et réflexion » par

rapport à l'enquête expérimentale⁶⁶. Il souligne la différence entre les jugements et intuitions de surface et les intuitions « robustes », qui seraient les vraies intuitions qui définissent nos concepts après qu'on a corrigé les biais, le manque d'attention et l'interférence pragmatique. Le philosophe serait mieux placé que le psychologue expérimental pour faire ressortir comment les gens raisonnent vraiment, parce qu'il faut un engagement actif, dialectique, rhétorique et émotionnel dans une conversation. D'après Kauppinen, le questionnaire en réponse à un scénario ou une petite histoire n'est qu'une approche quasi-observationnelle à la pensée des gens, incapable d'accéder aux subtilités. Il faut toucher dans le vif : on va avoir besoin au moins d'une bonne expertise dialectique, sinon d'un transfert émotionnel.

Ce genre de critique montre la conception caricaturale que les philosophes ont de la méthode expérimentale en psychologie. De deux choses l'une : soit on se réfère aux limitations d'expériences déterminées, soit on avance une critique valide en principe. Kauppinen se place carrément dans le deuxième champ, voulant défendre en général l'approche de la philosophie du canapé. Mais l'accusation selon laquelle la psychologie serait en général limitée à des méthodes quasi-observationnelles dans l'étude des concepts est caricaturale. Pour autant qu'on puisse rester dans un cadre rigoureux, la tendance actuelle en psychologie est à la pluralité méthodologique radicale : on va du questionnaire à l'observation, l'évaluation du comportement ou d'un dialogue par plusieurs évaluateurs qui atteignent un niveau d'accord acceptable entre eux, la confrontation d'hypothèses avec des données d'imagerie cérébrale, etc. S'il y a une possibilité d'évaluation objective (au moins dans le sens où elle serait partagée par plusieurs évaluateurs), les méthodes participatives et « chaudes » comme un dialogue avec le participant sont largement mobilisées, en synergie systématique avec toutes les autres. Kauppinen serait contraint de soutenir qu'il faut nécessairement le savoir-faire ineffable du philosophe, idée clairement non démontrable et pétition de principe. Il est clair qu'étudier les concepts et le raisonnement des gens est difficile et requiert un effort pluridisciplinaire. Il n'y a pas de raisons probantes pour soutenir que seule la philosophie peut atteindre ce but.

La nécessité de la philosophie expérimentale dans le cadre du fiabilisme

Peut-être tout ce qu'on vient de dire est-il basé sur une équivoque, car les philosophes n'auraient jamais dû s'intéresser aux concepts des gens : elle est bien la position des fiabilistes sur le rôle des

⁶⁶ A. Kauppinen, "THE RISE AND FALL OF EXPERIMENTAL PHILOSOPHY," *Philosophical Explorations* 10, no. 2 (2007).

intuitions⁶⁷. Pourquoi devrait-on prendre très au sérieux certaines intuitions, par exemple l'intuition qu'on ne sait pas qu'on n'est pas des cerveaux dans une cuve ? D'après le populiste épistémique, parce que tout le monde le pense : commencer par là a une bonne plausibilité stratégique. Pour le populiste, l'argument stratégique n'est pas que « si tout le monde le pense, alors cela doit être vrai », car un populiste peut aussi soutenir une théorie de l'erreur généralisée à la Mackie, et rejeter la validité de toutes les intuitions du sens commun. L'essentiel n'est pas la vérité des intuitions mais le fait même que tout le monde y soit déjà engagé, ce qui donne un point de départ pour un discours qui cherche à ré-articuler la dimension normative. Le populisme est avant tout une thèse méthodologique : une théorie normative ne pourrait que *commencer* par le sens commun.

Rien de tout cela pour le fiabiliste : la méta-justification du recours aux intuitions des philosophes (oui, uniquement celles des philosophes) est tout simplement qu'elles sont censées être fiables. À partir de cette position favorable, le philosophe considère les intuitions et les concepts des gens en réformateur : si les gens ont une vision de la connaissance incompatible avec celle des philosophes – par exemple, s'il pensent qu'on sait qu'on n'est pas un cerveau dans une cuve – c'est qu'ils devraient changer leur concept de connaissance. Autrement dit, ils devraient commencer à s'intéresser vraiment à *la* connaissance, étant donné que, dans cette conception, les intuitions des philosophes ne sont que des croyances fiables portant sur des phénomènes réels du monde. Les gens n'auraient pas une conception alternative, mais une conception fautive du phénomène concerné.

Maintenant, il faut bien sûr se poser la question suivante : (notre deuxième question, voir ci-dessus) pourquoi les philosophes auraient-ils des croyances intuitives plus fiables que les autres ? Mais, avant, il faut analyser la réponse des fiabilistes à la première question (pourquoi faire confiance à des intuitions, tout court ?), car l'idée de donner une place importante à des croyances intuitives censées être vraies ne va pas de soi : elle s'inscrit dans un débat complexe en théorie de la connaissance contemporaine.

On n'a pas vocation à ouvrir ce débat ici : il suffit de remarquer que l'existence de facultés cognitives sub-conscientes (donc intuitives) qui produiraient des croyances vraies et justifiées est acceptée par plusieurs théoriciens. En théorie de la justification, elle permet d'éviter le dilemme entre une régression à l'infini (dans la chaîne de justification d'une croyance explicite donnée) et une position cohérentiste.

Si on se pose la question de savoir d'où vient la justification d'une croyance explicite entretenue par un sujet, une réponse possible est qu'elle provient d'autres croyances explicites que le sujet

⁶⁷ E. Sosa, "Experimental philosophy and philosophical intuition," *Philosophical Studies* 132, no. 1 (2007), T. Williamson, "Philosophical 'Intuitions' and Scepticism about Judgement," *Dialectica* 58, no. 1 (2004).

entretient. Cette option est attractive parce qu'elle rend possible, du moins en principe, une pratique de justification très satisfaisante d'un point de vue dialogique : citer une série de croyances explicites à l'appui de la croyance de départ. « Je crois qu'il pleut », pourquoi cette croyance est-elle justifiée ? « Parce que j'ai vu la météo ». Dans un contexte dialectique, à chaque demande de justification il y aura une réponse précise. Cependant, étant donné qu'un esprit humain fini ne peut entretenir une infinité de croyances, il n'est pas difficile de se rendre compte que la chaîne de justification doit forcément tourner en rond⁶⁸, ce qui signifie qu'aucune croyance explicite n'est justifiée. À moins qu'on n'accepte une théorie cohérentiste de la justification, qui considère comme justifiée n'importe quelle croyance explicite dans un réseau cohérent d'autres croyances explicites. Mais cette deuxième solution est normalement rejetée, car il peut y avoir plusieurs ensembles de croyances différents et également cohérents, alors que la réalité est censée être unique : cela détache de façon inacceptable la question de la justification d'une croyance de sa vérité (supposée être unique et objective). Si la vérité est unique, alors que les ensembles cohérents de croyances justifiées sont très nombreux, c'est que les notions de vérité et justification ont un lien très relâché. Or, normalement la justification d'une croyance devrait nous donner des preuves que la croyance est vraie. Le cohérentisme est rejeté parce qu'il produit ce découplage inacceptable.

L'idée de croyance intuitive (on ne saurait pas comment la justifier explicitement) mais justifiée – parce que produite par des facultés cognitives fiables – est en mesure de résoudre le problème. Si une croyance intuitive (qu'on ne saurait justifier par d'autres croyances explicites) peut être justifiée, cela assure un ancrage dans le monde pour le réseau des croyances explicites et, pour ainsi dire, alimente de justification (comme s'il s'agissait d'essence) le système. Cela explique la possibilité d'un transfert de justification d'une croyance à l'autre (croyance *intuitive* p → croyance explicite q → croyance explicite r) et reste compatible avec la capacité d'un réseau cohérent de produire davantage de justification : par exemple, Sosa a soutenu qu'un réseau cohérent construit autour d'une croyance intuitive fiable est en mesure de lui conférer un « surplus » de justification. Pour résumer, l'acceptation d'une base de croyances intuitives produites par des facultés cognitives fiables semble en mesure de sauver la mise. Grâce à cet « ancrage » du système des croyances, on sauvegarde en même temps la pratique de la justification explicite (la croyance p soutient la croyance q) et la valeur épistémique de la cohérence. Telle est la stratégie de l'épistémologie de la vertu en philosophie de la connaissance contemporaine⁶⁹.

⁶⁸ Car, selon ce principe, la justification de chaque croyance vient d'une *autre* croyance : la chaîne ne peut pas simplement s'arrêter à un moment donné.

⁶⁹ E. Sosa, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge University Press, 1991).

Le cadre épistémologique des fiabilistes nous montre qu'on n'a pas le droit d'appliquer une suspicion généralisée à la dimension de la cognition intuitive : sous peine d'un scepticisme généralisé et de l'impossibilité d'attribuer une justification à nos croyances, on ne peut pas assumer une attitude juridique face à la justification, en demandant toujours une articulation explicite des raisons. Parfois, il faut faire confiance à l'intuition.

Dans ce contexte général, revenons à la deuxième question. Pourquoi les intuitions des philosophes ? Parce qu'il faut « faire confiance à l'intuition des experts ». L'expertise en question ne concerne pas « les concepts des gens » comme dans le cas du populiste, mais, tout simplement, le sujet dont il est question, c'est-à-dire la connaissance dans le cas des intuitions sceptiques, ou la morale en éthique. Le fiabiliste défend un parallèle entre l'expertise des philosophes et l'expertise standard dans n'importe quel domaine. La stratégie est très ingénieuse de la part des philosophes, car l'importance de l'intuition dans la cognition des experts est beaucoup plus qu'un lieu commun : elle a été largement démontrée par la psychologie⁷⁰. Très souvent, les experts ne savent pas justifier clairement ou pleinement leurs conseils ou diagnostics. Une grande partie de l'expertise est intuitive. Tout comme il faut faire confiance aux intuitions d'un ingénieur, d'un mathématicien, d'un physicien ou d'un bon joueur d'échec, il faudrait faire confiance aux intuitions morales et épistémiques des philosophes, en tant qu'experts de la connaissance et de la morale.

L'enseignement des experts est en général très réformateur par rapport aux intuitions des gens : il suffit de penser à l'effort nécessaire pour contrer les intuitions spontanées sur le fonctionnement des systèmes physiques, ou pour apprendre à envisager la structure d'un problème mathématique de façon correcte. Même si le gouffre entre les intuitions des philosophes et celles des gens est certainement moins important que celui qui sépare les bonnes intuitions physiques des prédictions spontanées sur la trajectoire d'un projectile, le philosophe serait également animé par une intention réformatrice. Quoi qu'ils en pensent déjà, il s'agit d'affirmer que les gens *devraient arriver à voir* qu'on ne peut attribuer de la connaissance dans un cas de Gettier, qu'« on ne sait pas qu'on n'est pas un cerveau dans une cuve » ou qu'« il n'est pas possible de sacrifier un innocent pour sauver trois condamnés ».

Cette méta-justification de la pratique du recours aux intuitions est vraiment remarquable dans sa simplicité, mais son point faible réside dans la nécessité de montrer, sans pétition de principe, que

⁷⁰ H. L. Dreyfus et S. E. Dreyfus, "The Ethical Implications of the Five-Stage Skill-Acquisition Model," *Bulletin of Science, Technology & Society* 24, no. 3 (2004), G. Klein, "The Power of Intuition," *Currency-Doubleday, New York, NY* (2003), E. Kurz-Milcke et G. Gigerenzer, *Experts in Science and Society* (Springer, 2004).

les philosophes peuvent être considérés comme des experts en philosophie de la connaissance ou en morale *tout court*, plutôt que des gens qui proposent *leur* théorie morale et *leur* théorie de la connaissance.

Comme le remarquent Bishop et Trout, normalement l'expertise, même la plus pointue, est associée à des critères d'évaluation indépendantes, publiques et facilement reconnaissables. Certes, à un moment donné il faut uniquement accepter les intuitions (qui peuvent nous paraître ésotériques) de l'expert, mais en général les experts en informatique sont ceux qui développent des logiciels qui marchent bien ; les ingénieurs font des ponts qui tiennent bon ; les mathématiciens démontrent des théorèmes qui ont beaucoup d'applications pratiques en cryptographie, ou dans la modélisation de phénomènes de tout genre ; les champions d'échecs gagnent des matches⁷¹. Le problème pour les philosophes est qu'il n'y a pas l'équivalent d'un critère indépendant d'expertise. Ils réfléchissent beaucoup sur la morale, mais est-ce qu'ils sont meilleurs que les autres d'un point de vue moral ? Est-ce que les institutions nées à partir d'engagements issus d'une réflexion philosophique sont meilleures que les autres ? Un expert n'est pas simplement quelqu'un qui a des opinions réfléchies sur une question, il s'agit de quelqu'un qui maîtrise, qui a du succès et des effets pratiques.

Sans une pétition de principe et une invitation à partager ses propres intuitions, comme si elles étaient évidentes et facilement accessibles à tout le monde (stratégie du populiste), le philosophe se coupe l'herbe sous les pieds et n'a aucune façon indépendante de montrer la pertinence de sa prétendue expertise. Au moins, le populiste part du constat d'un partage d'engagements : il s'adresse aux gens en faisant remarquer que, de toute façon, nous sommes tous déjà engagés à rejeter la possibilité de sacrifier un innocent pour sauver trois condamnés : l'universalité de cet engagement constitue un point de départ plausible.

Dans le cadre fiabiliste, ce dont les philosophes auraient besoin est la figure du philosophe-sage (ou du philosophe-saint) qui affiche une supériorité morale évidente dans son action et au jugement duquel on fait ensuite confiance. Cette proposition a deux types de problèmes. Elle est problématique en principe, parce que, s'il est certainement plausible de faire confiance et écouter le jugement d'un sage, il semble exagéré et contraire aux principes d'une négociation collective de donner à ses intuitions une place fondationnelle aussi nette. On rappelle que dans la pratique de la philosophie morale on donne à la plupart des intuitions du sens commun un rôle semblable aux observations en physique. Le conseil d'un sage mérite notre attention, mais la morale est un facteur trop important de l'interaction sociale et collective pour ne pas soumettre à la délibération collective

⁷¹ Michael A. Bishop et J. D. Trout, *Epistemology and the psychology of human judgment* (New York: Oxford University Press, 2005).

les verdicts intuitifs et péremptoires d'un sage ou d'un saint. L'objection ne consiste pas à nier toute expertise au sage, mais à nier une expertise suffisante pour qu'une société accepte ses verdicts.

Il y aurait aussi d'autres problèmes comme le désaccord entre différents sages: ce problème serait plus grave que dans le cas du populisme. Le populisme veut trouver des intuitions universelles, mais il peut réagir au désaccord de façon constructive, cherchant à expliquer au niveau psychologique la confusion conceptuelle qui a amené certains à l'erreur. Ces tentatives d'étiologie de l'erreur sont constructives parce qu'elles sont falsifiables de manière empirique. Les gens qui pensent qu'on pourrait jeter une personne d'un pont pour arrêter un train qui irait tuer 3 personnes (11% de la population)⁷² font peut-être une confusion conceptuelle entre le moyen et la conséquence d'une action. S'ils se rendaient compte qu'il s'agit de tuer une personne *comme moyen pour atteindre le but*, ils changeraient d'idée. Cette hypothèse est vérifiable : on peut tester si les personnes font cette confusion (moyen/conséquence) et voir comment leur jugement change après qu'on essaie de corriger la faute. Si le test empirique est négatif, on peut proposer une autre hypothèse explicative de l'erreur, et ainsi de suite.

Par contre, dans le cas du fiabiliste, à défaut de critères objectifs d'expertise, on doit faire confiance à la perception du sage. Suivant cette méthodologie, *par construction*, il n'y a pas de moyen de départager deux sages qui inspirent une grande admiration morale et qui seraient en désaccord sur une intuition fondamentale. Au-delà de ces problèmes de principe, l'idéal du sage s'accorde mal avec le contexte de la philosophie analytique du vingtième siècle, faite par des professeurs modestes et peu connus, dans des départements de philosophie isolés. Pour ces raisons, il me semble que la méta-justification fiabiliste est inacceptable et inapplicable au cas de disciplines normatives telles que l'éthique et la philosophie de la connaissance. Elle est inacceptable en soi, qu'on la formule et développe pour défendre la philosophie du canapé ou pour d'autres raisons. Par contre, si le populisme normatif n'est pas en mesure de sauver le philosophe qui se désintéresse des données empiriques, il est une stratégie intéressante qui peut être reformulée dans un contexte empirique, car des structures universelles de la cognition morale sont sans doute un point de départ intéressant pour une théorie normative.

La psychologie morale est l'étude des engagements dynamiques du sens commun : des intuitions à une jurisprudence naïve

⁷² M. D. Hauser et al., "A dissociation between moral judgments and justifications," *Mind and Language* 22 (2007).

La stratégie du populisme normatif consistant à partir d'engagements universels est intéressante, mais pour qu'elle puisse être reformulée dans le cadre d'une philosophie expérimentale, il faut travailler avec une conception beaucoup plus raffinée du sens commun et de la façon dont la philosophie devrait s'en servir. Pour l'instant, on a montré l'incapacité de la philosophie du canapé de découvrir quelles sont les bonnes intuitions du sens commun. Maintenant, il s'agit de montrer que le modèle où des intuitions joueraient un rôle « fondationnel » et « quasi-observationnel » est à dépasser en soi, même si on pouvait étudier ces intuitions de manière adéquate.

En effet, après avoir décidé de prendre au sérieux le sens commun, encore faut-il voir à quelles structures (situées à quel niveau d'analyse) psychologiques communes on s'intéresse. On va montrer que le format de l'intuition n'est pas suffisant pour représenter « la voix du sens commun » et ses engagements théoriques. L'erreur d'un philosophe qui se concentre sur une liste d'intuitions est tout à fait semblable à l'erreur d'un juriste qui se concentre sur une série de verdicts, sans chercher à reconstituer l'« esprit » de la loi. Plus exactement, l'erreur consiste à penser que les verdicts représentent une base suffisante pour comprendre l'« esprit » de la loi. Bien sûr, les verdicts sont une manifestation typique de l'esprit de la loi, mais pas la seule. Il faut également prendre en considération l'intégralité ou au moins une grande partie du cadre législatif. Autrement dit, il faut un point de vue holiste. De la même façon, les intuitions sont une manifestation importante des engagements théoriques du sens commun, même le philosophe doit s'intéresser à la cognition morale et sociale dans sa globalité, pour dégager « la voix » du sens commun.

Les intuitions sont un produit cognitif complexe

Les intuitions sont le résultat, le produit final de processus cognitifs et inférentiels complexes : si jamais il y a des intuitions universelles, *c'est avant tout grâce à l'activation de stratégies inférentielles, principes implicites ou stratégies perceptives, et non pas le contraire*. Une intuition très enracinée est basique au sens où qui l'éprouve trouve inconcevable la possibilité de penser le contraire. Mais elle n'est pas basique au sens cognitif : pour juger qu'il est absurde de jeter d'un pont une personne pour en sauver trois, il faut passer par des étapes complexes de catégorisation d'une action intentionnelle, de recours à des moyens qui paraissent justifiés ou non, d'évaluation de conséquences. La sensation d'immédiateté phénoménologique ne saurait écarter ces étapes de catégorisation profonde.

Cela reste vrai même dans un cadre empiriste et non-intellectualiste. Pour comprendre que les intuitions sont un produit cognitif complexe, on n'a pas besoin de penser qu'elles sont produites par la mobilisation des principes plus basiques d'une grammaire innée et profonde. Par exemple, même

si le fait de juger quelque chose comme immoral se réduisait au fait d'associer une émotion à la situation concernée, cette association ne serait pas simple et immédiate⁷³. Il s'agirait de voir quel niveau de description et de catégorisation implicite des caractéristiques de la situation déclenche la réaction émotionnelle. Même un simple processus de *pattern recognition* est en réalité axé et sensible à certains indices plutôt qu'à d'autres. Le rejet émotionnel déclenché par une liste de situations ne peut être un « fait cognitif brut ».

D'un côté, aucune intuition normative du sens commun n'est génétiquement canalisée de façon fiable et systématique, comme le nombre des doigts de la main. L'histoire, l'anthropologie et la psychologie morale contemporaines ont montré qu'il y a toujours des exceptions, fussent-elles très minoritaires : des personnes saines qui ne partagent pas certaines intuitions qui nous paraissent évidentes (voir ci-dessous). De l'autre côté, aucune intuition n'est fixée une fois pour toutes par une espèce d'*imprinting* culturel, devenant ensuite (à partir de ce moment) un fait cognitif brut pour la personne concernée. Les intuitions sont stabilisées et retravaillées par des processus et des forces psychologiques qui expliquent leur évolution. Ce cadre reste valide même si, en pratique (dans un ensemble très vaste de situations et histoires de vie), il est très difficile de changer ou de ne pas développer certaines intuitions déterminées.

Mais si une intuition est un produit cognitif complexe qui peut être influencé à plusieurs phases intermédiaires d'élaboration, on voit bien que les engagements les plus basiques du sens commun ne sauraient être identifiés aux intuitions, mais aux principes, aux modes de gestions et d'élaboration qui les produisent. En d'autres termes, il faut renverser complètement l'objectif de développer une théorie ou une jurisprudence morale théorique qui respecte nos intuitions et, au contraire, découvrir la jurisprudence implicite qui les anime. C'est la jurisprudence naïve ou (l'esprit de la loi) qui gère l'évolution dynamique de nos engagements superficiels et qui représente la vraie dimension des engagements profonds du sens commun.

L'étude des engagements dynamiques du sens commun

Nous avons déjà montré qu'une référence à une intuition ne peut être qu'une version elliptique d'affirmations contrefactuelles beaucoup plus coûteuses sur la façon dont les gens réagiraient et ce qu'ils diraient face à d'autres stimuli : par exemple, une théorie basée entièrement sur des intuitions

⁷³ J. Haidt, "The emotional dog and its rational tail: A social intuitionist approach to moral judgment," *Psychological Review* 108, no. 4 (2001), Jesse Prinz, *The Emotional Construction of Morals* (Oxford: Oxford University Press, 2007).

anti-utilitaristes présuppose clairement que les gens ne réviseraient pas ces mêmes intuitions du jour au lendemain face à la première critique. L'engagement du sens commun doit être par essence dynamique : la possibilité qu'en effet une liste d'intuitions reste fixe quoi qu'il arrive (comme les philosophes l'ont toujours pensé) n'est qu'un cas de figure particulier qui n'invalide pas le principe. Cela voudrait dire que la jurisprudence intuitive appliquée à des cas simples donne toujours les mêmes réponses. Mais même dans ce cas de figure, il serait incorrect d'identifier la position du sens commun à cet ensemble d'intuitions et non pas à la jurisprudence intuitive qui les a produites. Le problème est qu'à partir d'un ensemble limité d'intuitions « statiques » il est impossible de prédire comment les gens sont implicitement engagés à raisonner face à des nouveaux cas : ce serait comme prédire des nouvelles délibérations juridiques uniquement à partir de précédents, non grâce (aussi) à la connaissance de la loi.

Mais étudier les intuitions n'est-il pas la seule façon de dériver la jurisprudence implicite ? Oui, bien sûr. La compréhension de la jurisprudence naïve passe nécessairement par une étude empirique systématique des jugements des gens. Le problème est plutôt l'usage que les philosophes en font. Le travail de tout philosophe qui part du sens commun n'est jamais une simple reconstruction de ce dernier : le but est toujours celui d'offrir une thèse informative et de modifier notre schéma conceptuel, même si ce n'est que de façon marginale et si le critère du succès de la théorie est celui de produire la moindre modification possible. S'il ne fallait produire strictement aucune modification, le travail du philosophe serait inutile, par principe. Dans un cadre populiste, l'autorité de la modification proposée est justifiée par le fait que la théorie philosophique ne fait que rendre cohérente la position du sens commun : ce n'est que du sens commun explicité, corrigé et purifié de paradoxes et contradictions.

C'est pour cela que les gens devraient accepter la théorie, parce que c'est leur position corrigée. Mais un aspect essentiel de cette méta-justification de la pratique philosophique est que, en amont du processus, la position du sens commun ait été correctement représentée. Sinon la procédure est fallacieuse. S'il est vrai que je désire profondément visiter un pays comme la Thaïlande, alors il s'en suit que, si ce n'est pas possible d'aller en Thaïlande pour des raisons pratiques, je devrais être motivé à visiter le Vietnam, qui est un pays très semblable (supposons que ce soit vrai) et où je peux trouver ce que je cherchais. Un ami pourrait me tenir ce discours. Mais le raisonnement devient fallacieux si l'engagement de départ n'était pas représenté correctement : peut-être, si je veux aller en Thaïlande c'est bien sûr que j'aime le pays, mais surtout que j'ai de la famille là-bas. En d'autres termes, il faut une représentation *profonde* et *complète* de l'engagement de départ pour avoir la justification de proposer une mise à jour, ce qui est l'ambition du philosophe.

Si tout ce que le sens commun avait à dire se concentrait dans une liste d'intuitions accessibles, alors le sens commun ne serait pas en mesure de refuser la dernière mise à jour proposée par le philosophe. Mais, sans une compréhension profonde de la jurisprudence naïve, on ne peut être sûr que les intuitions ne vont pas changer. Prenons l'exemple d'une société du passé qui considère normal de tuer des animaux et où personne n'a des intuitions de végétarien ou de défenseur des droits des animaux. Un philosophe moral qui n'avait accès qu'au sens commun de cette société pourrait considérer « tuer les animaux est licite » comme une intuition de base. Imaginons que la culture en question entre en contact avec une culture scientifiquement plus avancée, qui possède des connaissances sur la neurobiologie des animaux et sur le fait qu'ils peuvent souffrir, et surtout souffrir d'une façon très semblable à celle de l'homme. La culture de la première société pourrait en être immédiatement influencée et personne ne serait plus sûr de ses intuitions. Une explication possible du changement rapide est que l'intuition de départ s'explique par le manque de vraie compréhension de la possibilité que les animaux souffrent comme les hommes : si cela est correct, alors cette société n'a pas seulement changé d'opinion, elle n'a été *jamais* engagée sur le caractère licite de la suppression des animaux de la manière *profonde* et *dynamique* qui intéresse le philosophe.

Même dans le cas où certaines intuitions seraient dynamiquement constantes, c'est-à-dire un vrai produit constant de la psychologie profonde humaine, s'arrêter au niveau d'une liste d'intuitions décisives pourrait créer un problème de sous-détermination qui invalide la stratégie du philosophe. Un noyau d'intuitions très profondes comme « ne pas jeter une personne du pont pour en sauver trois » pourrait être compatible avec plusieurs jurisprudences naïves possibles. Mais ces reconstructions différentes de la jurisprudence naïve pourraient ne pas être équivalentes sur des nouveaux cas inusuels. Ainsi, un philosophe qui arriverait à une hypothèse sur la position profonde du sens commun à partir de l'analyse de certaines intuitions cruciales pourrait proposer une mise à jour qui est incompatible avec la vraie jurisprudence profonde, alors que la reconstruction arrivait à prédire correctement les cas de départ. Pour résumer, si je ne sais pas comment tu raisones, je peux, au mieux, connaître tes propos habituels mais non ce que tu dirais dans une situation non commune. Ainsi, le philosophe ne pourra être sûr que sa mise à jour du schéma conceptuel sera acceptée et n'ira pas toucher des engagements cachés du sens commun.

Nous croyons que le problème est clair d'un point de vue théorique. Cependant, nous pourrions être accusés d'être trop sceptiques sans raison : certes, si les mécanismes profonds qui engendrent les intuitions réservaient des surprises incroyables, alors s'arrêter aux intuitions ne serait pas suffisant,

mais apparemment il y a un corpus d'intuitions bien connues et stables. Par exemple, les gens ne semblent pas être des utilitaristes radicaux, et personne n'a découvert des argumentaires qui puissent les convertir rapidement à l'utilitarisme, montrant que les intuitions anti-utilitaristes étaient superficielles. Il semble qu'il faille être sceptique pour douter que la jurisprudence morale naïve est anti-utilitariste.

Peut-être l'anti-utilitarisme est-il très enraciné dans la psychologie des gens, mais, il ne faut pas oublier que des études conduites sur Internet avec un échantillon de 5000 personnes dans 120 pays différents ont montré que 11% des gens ont l'intuition utilitariste de jeter la personne du pont pour en sauver trois. Ce résultat persiste à tous les niveaux si on analyse les sous-groupes par nationalité, sexe ou âge⁷⁴. Une personne sur dix ne constitue pas une exception dramatique pour des fins pratiques, mais le philosophe populiste est à la recherche d'engagements universels : il ne peut pas se contenter de laisser passer comme cela qu'une minorité de la population est amoraliste ou légèrement psychopathe. Il est contraint d'interpréter l'exception comme une erreur de mesure ou une confusion, et avancer une théorie de l'erreur. Mais ce qu'il est nécessaire de découvrir pour avancer une théorie de l'erreur qui tienne est justement la façon de raisonner des gens, la jurisprudence intuitive et ce qui tourne mal dans son application par le 11% des gens utilitaristes (à supposer qu'il y ait quelque chose qui tourne mal). Ainsi, à défaut d'un accord vraiment universel sur les intuitions de base, le populiste a besoin de découvrir la jurisprudence naïve pour essayer d'expliquer l'erreur de l'amoraliste.

Section II - Qu'est une théorie naïve en psychologie ?

Jusqu'ici, on a vu que l'approche traditionnelle en philosophie ne pose pas correctement la question de l'étude du sens commun. Le projet d'accéder aux « intuitions du sens commun » est mal défini. La notion de sens commun est trop vague et, par principe, un corpus limité d'intuitions n'est pas en mesure de lui rendre justice. Essayons maintenant de voir comment la question de découvrir la position du sens commun a été affrontée en psychologie.

Depuis des décennies, les psychologues parlent de « théories naïves », par exemple en ce qui concerne la physique ou la biologie populaires, mais aussi par rapport à la psychologie naïve qui mobilise des états mentaux tels que croyances et désirs. Récemment, il y a eu une explosion de théories naïves sur presque tout (action, intentionnalité, connaissance, etc.), mais plusieurs

⁷⁴ Hauser et al., "A dissociation between moral judgments and justifications."

démarches théoriques emploient le même label sans qu'il y ait encore un vocabulaire accepté pour marquer les différences. Nous chercherons à clarifier conceptuellement ce que cela pourrait signifier que le sens commun cautionne implicitement une théorie. Les distinctions théoriques générales élaborées seront décisives dans la discussion spécifique de la théorie naïve du caractère, qui va nous occuper dans le prochain chapitre.

Dans l'usage courant, très peu discipliné, presque toute étude portant sur la cognition ordinaire dans un certain domaine peut se retrouver associée au label « théorie naïve », par un biais ou par un autre. La tentation de tirer des conclusions sur la théorie naïve du sens commun – alors qu'on est tout simplement en train d'étudier le fonctionnement de la cognition ordinaire – est justifiée par deux raisons.

La première raison est que tout aspect de la cognition ordinaire – même si spécifique et isolé – semble indiquer « l'adhésion » à certaines positions théoriques, au moins dans le sens que ces positions s'y accordent mieux que d'autres et semblent presque être « suggérées ». Par exemple, si on montre que les gens pensent qu'on peut être responsable d'un homicide dans un monde parfaitement déterministe, à la rigueur on a uniquement étudié les attributions de responsabilité que les gens font dans une classe de cas. Mais cela semble appuyer l'idée (plus générale) selon laquelle les gens seraient compatibilistes⁷⁵. Il s'agit là de la pratique de « lire » les engagements théoriques implicites des gens directement à partir d'aspects particuliers du fonctionnement de la cognition. Bien sûr, l'attribution est plus solide et justifiée s'il y a plusieurs aspects dans la psychologie populaire qui vont dans la direction souhaitée. Pour être charitable, on peut croire que les auteurs le savent bien et qu'ils présupposent qu'il doit y avoir beaucoup plus de manifestations de l'attitude dont il est question.

La deuxième raison pour parler de théories implicites sur un domaine est l'existence fréquente de représentations explicites et répandues portant sur le domaine étudié. Par exemple, des croyances répandues et (qui paraissent) spontanées sur la biologie et les espèces⁷⁶. En ce sens, en étudiant la cognition ordinaire, on s'aperçoit qu'il n'y a pas que des engagements théoriques manifestés indirectement par l'attitude des gens, souvent le sens commun entretient certaines croyances de façon explicite.

⁷⁵ Le compatibilisme est la thèse selon laquelle le libre arbitre et la responsabilité morale sont compatibles avec le déterminisme. E. Nahmias et al., "Surveying Freedom: Folk Intuitions about free will and moral responsibility," *Philosophical Psychology* 18, no. 5 (2005).

⁷⁶ D. L. Medin et S. Atran, *Folkbiology* (Bradford Book, 1999).

Une fois comprises les raisons de l'usage relâché de la notion de théorie naïve, cherchons à élaborer une classification plus significative. En réalité, dans chaque domaine, la littérature psychologique associée au label de « théorie naïve » est typiquement occupée à étudier un ou plusieurs des aspects suivants de la cognition :

L'*architecture cognitive* et les mécanismes psychologiques (qui gèrent les processus psychologiques pertinents) du domaine concerné

Les *concepts naïfs* qui sont mobilisés dans ces processus psychologiques

Les *principes naïfs* qui sont vraiment codés et opérationnels dans le système (s'il y en a)

Le *fonctionnement* du domaine cognitif : l'ensemble d'inférences, attributions, prédictions, évaluations et interprétations qui mobilisent les concepts naïfs et qui sont typiquement produites dans les processus psychologiques qui constituent le domaine cognitif.

La *fonction* évolutive ou actuelle (s'il y a une différence) du domaine cognitif

Les *engagements ontologiques* du système cognitif (les entités et les processus dont l'existence est « postulée » par le système cognitif)

Des intuitions naïves déterminées

La *théorie naïve spontanée* : ce que les gens diraient explicitement – spontanément ou si on leur demande – sur les entités, propriétés ou processus concernés par le domaine cognitif.

Pour clarifier ces notions, il vaut mieux commencer par des exemples. La psychologie morale est un bon cas d'étude pour articuler ces différents niveaux, car elle est étudiée depuis tous ces points de vue. Commençons par la fin de notre liste. D'abord, on peut se limiter à étudier la façon dont les gens définiraient explicitement la moralité⁷⁷. Ensuite, comme on l'a vu, sous le label de « théorie naïve », il y a des études qui explorent les intuitions communes sur des cas spécifiques, comme le problème du « trolley » (en philosophie de la connaissance, les scénarios sceptiques ou de Gettier)⁷⁸. Les données de ce genre d'études peuvent être utilisées pour analyser les concepts moraux naïfs, tels que « juste » ou « moralement acceptable ».

Les jugements intuitifs sont le produit final de processus psychologiques d'inférence, interprétation, attribution, prédiction qui mobilisent les concepts moraux : une partie de la littérature psychologique cherche à reconstruire la nature et la portée de ces processus (*fonctionnement* du domaine cognitif), qui ont lieu dans notre architecture cognitive et neurologique. Il s'agit véritablement de cartographier le comportement du système cognitif : idéalement, il faudrait

⁷⁷ Stich est en train d'accumuler des données expérimentales sur les croyances spontanées concernant la morale, dans plusieurs cultures.

⁷⁸ J. Greene et J. Haidt, "How (and where) does moral judgment work?," *Trends in Cognitive Sciences* 6, no. 12 (2002), S. Nichols, S. Stich et J. Weinberg, "Metaskepticism: Meditations in Ethno-Epistemology," *The Sceptics* (2003).

associer un output à chaque état interne et input du système. Bien sûr, on n'est pas en mesure de faire cela d'une façon aussi rigoureuse, mais on peut apprendre beaucoup sur la façon typique dont les gens raisonnent et réagissent dans plusieurs contextes pertinents pour la morale. En ce sens, il est possible d'avancer dans cette démarche descriptive même si la compréhension de l'architecture cognitive et des mécanismes profonds est à la traîne. Tout comme on peut mémoriser les patterns de comportement d'un ami dans plusieurs contextes (travail, famille, sport) sans avoir de bonnes explications sur ses dynamiques profondes – ou comme on peut savoir par expérience comment réagira Windows dans une longue série de situations spécifiques, sans être un programmeur – on peut cartographier les réactions typiques à des tâches moralement pertinentes, sans disposer de théories profondes. Cette situation est typique de la cognition sociale. Des décennies de traditions expérimentales de grande qualité ont mis au jour des effets dont l'explication profonde ne fait pas consensus. On possède un savoir très vaste sous la forme suivante : dans la situation naturelle (ou dans le paradigme expérimental) qui a les caractéristiques x, y, z , les gens réagissent de façon c . Le défi expérimental est celui de montrer que l'effet se produit dans une situation définie de façon rigoureuse et reproductible. Par exemple, dans le paradigme de l'obéissance de Milgram, en moyenne deux tiers des participants obéissent (voir le chapitre sur la littérature situationniste). L'accumulation de données de ce format nous donne une image à haute définition du fonctionnement de la cognition ordinaire, alors que les explications profondes se font attendre. Pour lui donner un nom et la distinguer d'autres démarches scientifiques, on appellera cette cartographie détaillée du comportement de la cognition ordinaire dans un domaine étude du « fonctionnement » de la cognition ordinaire dans ce domaine.

Enfin, pour compléter la clarification de notre liste grâce à l'exemple de la cognition morale, il ne faut pas oublier l'étude des principes innés (s'il y en a) actifs dans l'élaboration de l'information morale à la base de nos jugements – une grammaire morale profonde – et l'étude de la fonction évolutive de la cognition morale⁷⁹.

Quelle (ou quelles) notion intéressante de « théorie naïve » peut-on retrouver à partir de tous ces niveaux ? On le sait, les définitions sont libres. Tant qu'on ne crée pas de confusion, on peut continuer à parler de théorie naïve en plusieurs sens relâchés : ainsi, on peut appeler « théorie naïve » du sens commun une série d'opinions explicites très répandues chez les gens (théorie spontanée). Par ailleurs, si on découvre certains schémas dans le fonctionnement de la cognition

⁷⁹ R. Joyce, *The Evolution of Morality* (MIT Press, 2006), J. Mikhail, "Universal moral grammar: theory, evidence and the future," *Trends in Cognitive Sciences* 11, no. 4 (2007).

ordinaire – par exemple, on dresse une longue liste de situations dans lesquelles les gens attribuent facilement des traits de caractère ou autres dispositions générales – on pourra dire que les gens ont une théorie naïve « dispositionniste ». Cette façon de procéder est compréhensible, mais on va défendre l'idée selon laquelle il y a un seul sens profond et intéressant de « théorie naïve » (aussi bien pour les psychologues que pour les philosophes) : les engagements ontologiques du sens commun. Il s'agit des entités et processus dont l'existence est postulée par le sens commun. Chacun des autres sens « locaux » et « relâchés » de théorie naïve est très intéressant en soi-même. Par exemple, découvrir les croyances explicites spontanées des gens a un intérêt évident en anthropologie cognitive et sociologie, et cartographier le fonctionnement de la cognition ordinaire est évidemment essentiel en psychologie. L'importance de ces démarches n'a pas besoin d'être justifiée. Mais, en philosophie, lorsqu'on cherche à « partir du sens commun » dans le cadre d'une stratégie populiste, on devrait s'intéresser uniquement aux engagements ontologiques.

La notion d'engagement ontologique a été rendue populaire par Quine⁸⁰, relativement aux théories scientifiques : depuis, elle est restée une notion cruciale en philosophie des sciences. D'après Quine, malgré la prolifération de textes scientifiques facilement accessibles, comprendre quel genre d'entités et processus sont véritablement postulés par une théorie est une opération complexe et « holiste ». Il faut prendre en considération la théorie dans sa globalité et réfléchir à la possibilité d'une reformulation et simplification. Peut-être, dans une version simplifiée, la théorie gardera le même pouvoir de prédiction mais postulera beaucoup moins d'entités problématiques. Le sens commun n'est pas du tout analogue à une théorie scientifique formalisée, mais il y a un aspect de l'analogie qui est fondamental : le caractère « holiste » de l'analyse nécessaire pour dégager les engagements ontologiques. Tout comme il faut prendre en considération une théorie scientifique dans sa globalité, il faut avoir une vision d'ensemble des différents niveaux d'analyse du sens commun qu'on a introduits, pour pouvoir en dégager les conséquences ontologiques. Pour cette raison, aucune tradition expérimentale spécifique ne peut correspondre à l'étude des engagements ontologiques du sens commun.

Nous allons clarifier ultérieurement ces distinctions et leur utilité heuristique en examinant deux confusions conceptuelles typiques dans la littérature : (1) extraire des engagements ontologiques directement d'aspects locaux du fonctionnement de la cognition ; (2) confondre une cartographie du

⁸⁰ W. V. Quine, *Le mot et la chose*, traduit par J. Dopp et P Gochet (Flammarion Paris, 1999).

fonctionnement de la cognition et une élucidation des engagements ontologiques⁸¹. Dans ces deux cas, les psychologues cèdent à la tentation de prendre des raccourcis et identifient un aspect local de la cognition ordinaire avec la « position » du sens commun, tout court.

Extraction d'engagements ontologiques d'aspects locaux du fonctionnement de la cognition

Nous avons montré que certains aspects du fonctionnement de la cognition ordinaire peuvent manifester ou suggérer des engagements théoriques dans le sens commun. Il est important d'articuler davantage cette notion de théorie implicite qui ne se manifesterait pas sous la forme d'un ensemble de croyances explicites de l'agent. La notion est générale et peut être appliquée à d'autres systèmes, non seulement à la cognition humaine.

Il vaut mieux partir d'un exemple. Prenons le cas d'un logiciel cryptographique complexe : ce logiciel est basé entièrement sur l'idée qu'un hacker ne serait pas en mesure d'effectuer des calculs trop complexes avec des nombres premiers. Cette idée n'est écrite nulle part dans le code, mais, si on ne la mobilise pas, la structure et la stratégie du logiciel ne semblent avoir aucun sens. Un hacker pourrait même découvrir que c'est bien là la stratégie des programmeurs. En ce sens, il est « implicite » dans le logiciel que « certains calculs avec des nombres premiers ne pourront pas être effectués par les hackers ». C'est un sens parfaitement descriptif, non mystérieux – et qui ne mobilise pas d'homuncules – de théorie implicite qui peut être appliqué à la cognition humaine. Pour peu qu'il y ait des engagements théoriques de ce genre, leur découverte ne peut être faite qu'à un niveau global, à partir d'une évaluation globale du domaine cognitif concerné. Il y a des raisons spécifiques à cela.

Dans une pluralité de domaines, la cognition humaine est une boîte à outils issue de l'évolution qui ne devrait pas être censée présupposer ni un ensemble parfaitement cohérent de postulats ni (tant que cela n'est pas prouvé) se ramener à un pot-pourri d'heuristiques complètement bricolé. Dans le premier cas, les psychologues devraient articuler un système théorique d'une cohérence exquise. D'après ce que nous savons de la façon dont l'évolution marche, cela est très peu plausible en ce qui concerne les différents domaines de la cognition sociale. Nos ancêtres ont été confrontés à une

⁸¹ Une troisième erreur – penser que les théories spontanées sont représentatives des vrais engagements ontologiques – est moins fréquente (ex. : on sait très bien que les croyances explicites sur la grammaire ne sont pas fiables), mais elle est très importante dans le cas de la théorie populaire des traits, et sera traitée dans le prochain chapitre.

grande quantité et variété de problèmes d'interaction sociale, avec des matrices de coûts-bénéfices cognitifs et pratiques extrêmement hétérogènes. L'évolution a tendance à bricoler, réutiliser les mêmes ressources et chercher des raccourcis ; il serait vain d'espérer que les principes qui régissent notre vision naïve du monde constituent une anthropologie cohérente et policée. Il est plus tentant de tomber dans l'excès contraire. Si la cognition sociale est un pot-pourri total, il n'y a pas de stratégie ou politique globale qui puisse émerger, et on aura seulement des heuristiques spécifiques pour des problèmes déterminés : on pourrait être programmé pour traiter d'une certaine façon les tensions avec des amis ou alliés, avoir une attitude complètement différente avec l'autorité, ou les enfants, etc.⁸²

L'idée même de lire des engagements ontologiques à partir de la cartographie du fonctionnement d'un domaine cognitif a du sens uniquement à un niveau global : si la cognition sociale est un pot-pourri totalement hétéroclite d'heuristiques, cela n'aurait pas de sens de lire des engagements ontologiques différents dans chacune d'entre elles. Imaginons que nous soyons programmés pour blâmer d'une façon déterminée les gens qu'on envie beaucoup, alors qu'on ne retrouve pas ailleurs ce traitement spécifique. Il n'y aurait aucun intérêt à affirmer qu'on a une théorie implicite du blâme des personnes enviées. Ce serait seulement une façon tautologique de répéter qu'on a une telle politique conceptuelle dans ce contexte. On ne ferait que réintroduire les faits sur le fonctionnement de la cognition qu'on a déjà découverts, sans toucher à la stratégie profonde. Dans le scénario du pot-pourri total, il n'y a tout simplement pas d'engagements ontologiques et aucun sens intéressant de théorie naïve.

Les psychologues sont souvent tentés par le scénario du pot-pourri total, spécialement pour les volets de la cognition sociale qui ne sont pas compris dans la psychologie des pro-attitudes (croyances, désirs, intentions). Les psychologues et les philosophes admettent l'existence d'une théorie de l'esprit universelle : dans toutes les cultures on retrouve des notions approximativement équivalentes à celles de désir et de croyances. Mais, mis à part ce volet, le reste de la cognition sociale semble être constitué par un ensemble hétérogène d'heuristiques, schémas, traits et stratégies de prédiction qu'il est difficile de catégoriser. Or, dans cette vision de la cognition sociale, on n'aurait pas le droit d'attribuer d'engagements ontologiques au sens commun.

Paradoxalement (mais de façon conséquente) cette attitude engendre l'effet contraire d'une surproduction d'hypothèses sur les engagements ontologiques, dans différents domaines. Des experts dans un nombre limité de tâches spécifiques essaient de généraliser leurs résultats à tout le domaine : par exemple, il a été suggéré que, vu qu'on attribue très facilement des dispositions dans

⁸² Gary Marcus, *Kluge: The Haphazard Construction of the Human Mind* (Houghton Mifflin Co, 2008).

la phase de formation des impressions, on serait des dispositionnalistes ; en général, nous nous promènerions dans notre monde social avec des « lunettes dispositionnelles », pour ainsi dire⁸³. La motivation des psychologues pour généraliser leurs résultats de cette façon est d'autant plus forte que le soutien empirique augmente : au fur et à mesure que la formation des impressions paraît se réduire de plus en plus à une « affaire dispositionniste », la pertinence de la pensée dispositionniste gagne en crédibilité et on commence à espérer que le dispositionnisme explique d'autres aspects de la cognition sociale.

Or, conclure à partir d'une attitude dispositionniste établie dans un bon nombre de tâches que les gens sont globalement dispositionnistes, au niveau de leurs engagements ontologiques, est certainement hâtif. Pire, c'est une instance de l'erreur consistant à penser que grimper sur les arbres nous rapproche d'une solution pour atteindre la lune. Bien sûr, si tous les processus impliqués dans la compréhension d'autrui au sens large mobilisaient des concepts dispositionnels, on pourrait en conclure que les gens portent des « lunettes dispositionnalistes ». Mais montrer qu'une partie de la cognition sociale est plus importante qu'on l'aurait imaginé ne nous rapproche pas de conclusions sur les engagements ontologiques. Découvrir qu'une entreprise de chaussures a une politique de gestion des stocks beaucoup plus élaborée et soignée que l'on ne le croyait ne fait rien pour prouver qu'il s'agit d'une entreprise de gestion des stocks (et non pas de chaussures). De la même façon, découvrir que, contrairement à ce qu'on n'aurait pu penser, dans un logiciel de traitement de texte, la moitié de la puissance de calcul requise est consacrée à l'interface graphique ne fait rien pour prouver qu'il s'agit d'un logiciel de gestion d'images. Une évaluation globale et *qualitative* de la cognition sociale liée à la compréhension d'autrui est nécessaire. On verra l'importance de l'aspect qualitatif en particulier en ce qui concerne la théorie implicite des traits de caractère, dans le prochain chapitre.

Je soupçonne les psychologues sociaux de généraliser si facilement à partir de leur domaine restreint – sans argumenter de façon adéquate – parce que, paradoxalement, ils ne pensent pas qu'il existe un *vrai* niveau stratégique global dans le pot-pourri de la cognition sociale. Ils sont sérieusement tentés par l'hypothèse du pot-pourri total. J'ai deux commentaires à faire sur cela. Premièrement, leur façon de s'exprimer (par exemple, lorsqu'ils affirment que les gens sont dispositionnistes) est incorrecte, même si en réalité ils n'avaient pas l'intention de mettre au jour des engagements ontologiques. De deux choses l'une : dans un certain domaine cognitif, un théoricien est soit en train de formuler quelques hypothèses (aussi limitées et timides qu'elles soient) sur les

⁸³ Lee Ross et Richard E. Nisbett, *The person and the situation: perspectives of social psychology*, McGraw-Hill series in social psychology (New York: McGraw-Hill, 1991).

engagements ontologiques soit en train de cartographier des aspects spécifiques du fonctionnement du domaine. Il n'y a pas d'autres possibilités. Affirmer que « les gens ont des lunettes dispositionnistes, au moins en ce qui concerne la formation des premières impressions » est seulement une façon de répéter la découverte que la formation d'impressions mobilise des concepts dispositionnels. Si on veut dire quelque chose de plus, comme cela est souvent le cas, cela ne peut que générer des confusions. L'idée d'engagements ontologiques « locaux », relativisés à un ensemble limité de tâches, n'a pas de sens.

Deuxièmement, d'un point de vue méthodologique, on devrait résister autant que possible à l'hypothèse du pot-pourri total, cela pour deux raisons. D'abord, pour peu qu'il y en ait vraiment, des engagements ontologiques (même extrêmement limités et vagues) implicites globalement dans la cognition sociale seraient ce qu'il y a de plus intéressant à apprendre d'un point de vue normatif et philosophique. En philosophie, comme on l'a vu, les engagements théoriques des gens ordinaires ont occupé une place centrale et très respectée dans la théorisation normative, le long d'un continuum qui va de la reconnaissance comme la source principale de normativité jusqu'au rejet total (mais soigneusement motivé). Si, au lieu d'être contraint – à la limite – à rejeter les positions du sens commun on devait reconnaître qu'il n'y en a pas, ce serait la perte d'une référence fondamentale pour les philosophes. Il est normal d'exiger de très bonnes raisons pour accepter cette situation.

S'il n'y a pas de stratégies générales et des engagements de fond, il est très peu probable que les philosophes aient beaucoup à apprendre du pot-pourri total de la cognition sociale. Par exemple, si pour des limitations évolutives et cognitives une certaine heuristique interprétative s'active lorsqu'on se souvient de vieux amis qu'on associe à des émotions spécifiques – et ne s'active pas dans des souvenirs qui ont une tonalité émotionnelle différente – cela n'est pas forcément dû à des raisons très profondes. Si on avait eu des heuristiques un peu ou très différentes, notre vie sociale aurait pu être aussi bonne et même meilleure. On peut même essayer d'acquérir des meilleures réflexes avec un entraînement cognitif adéquat. Les détails d'un pot-pourri d'heuristiques ne sont pas porteurs d'enseignements normatifs intéressants. Par contre, il y a une chance que des engagements ontologiques généraux soient là pour des motifs plus structuraux et soient porteurs d'une sagesse qu'il faut au moins prendre en considération, d'un point de vue philosophique.

Si l'on découvre comment un logiciel cryptographique traite certains cas ou sub-phases computationnelles, cela est très bien. Mais si on se rend compte qu'il implémente globalement une certaine stratégie cryptographique qui a un ensemble défini de postulats, on a atteint une compréhension descriptive plus profonde de ce qu'il fait et de la façon dont il fonctionne. Il en va

de même pour un domaine de la cognition sociale. L'hypothèse de l'existence d'engagements ontologiques devrait être écartée uniquement avec de très bonnes raisons.

La deuxième raison pour être sceptique quant au scénario du pot-pourri total est la suivante : contrairement à ce qu'on pourrait croire, dans un domaine cognitif, il ne faut pas postuler un ordre parfait et improbable pour pouvoir repérer des engagements ontologiques. Un système imparfait, un peu bricolé, fait de sous-systèmes recyclés et qui localement peut avoir des aspects dysfonctionnels (bref, un système cognitif issu de l'évolution) peut suivre quand même une stratégie de fond. S'il y a un niveau stratégique sous-jacent qui s'impose à l'analyse, on peut accepter facilement des contradictions locales. On a le droit de rester optimiste et de s'engager à fond dans l'effort d'une évaluation qualitative du fonctionnement d'un certain domaine de la cognition sociale, sans se décourager si un pattern intéressant ne s'impose pas depuis le début.

Confusion entre le fonctionnement d'un domaine cognitif et ses engagements ontologiques

Parfois le problème n'est pas d'ignorer la nécessité d'un point de vue global sur un domaine cognitif. Par exemple, en ce qui concerne l'explication et la compréhension du comportement d'autrui, au fil des décennies, la cartographie des attitudes attributives et explicatives a été faite avec un niveau de précision remarquable. On sait exactement dans quelles cultures et dans quelles situations expérimentales ou naturelles les gens ont tendance à utiliser certains concepts plutôt que d'autres (croyances, intentions, traits, désirs, plusieurs heuristiques, etc.). Ces données sont très informatives et souvent surprenantes. Plusieurs psychologues comprennent très bien qu'elles dressent un tableau global qui demande une vue d'ensemble. Toutefois, ils ne s'aperçoivent pas que ce tableau mérite aussi d'être interprété. Ils pensent que la description détaillée de l'usage des concepts ordinaires nous délivre automatiquement les engagements ontologiques du sens commun. Prenons l'exemple des efforts récents de Malle pour étudier les explications dans la psychologie populaire. Dans une série d'expériences remarquables, Malle a dressé un tableau assez complet des différents modes d'explication mobilisés par les gens dans des contextes ordinaires. Sa théorie opère des distinctions entre plusieurs types d'explications causales et intentionnelles, et montre quels concepts sont mobilisés dans chaque cas de figure.

Voici quelques exemples de cette cartographie détaillée, notamment par rapport aux différences dans l'usage des concepts de croyance, désir, trait, but, etc. :

Asymétrie désir/croyance : les explications intentionnelles mobilisent surtout les désirs de la cible : lorsqu'on explique le comportement de quelqu'un, on se concentre surtout sur ses désirs. « Il a fait

cela parce qu'il aime x, y, z ». Il faut une connaissance appropriée du sujet et la motivation d'être plus rigoureux pour que les sujets commencent à attribuer *aussi* des croyances (« Il a fait cela parce qu'il aime x et il croit y, z »).

Asymétrie explications intentionnelles/autres explications : dans le choix entre des explications intentionnelles (c'est-à-dire des explications qui mobilisent les notions de désir et croyance) et des explications qui se réfèrent à la personnalité (traits, émotions, buts), les premières sont choisies pour des raisons de gestion d'une réputation sociale : il semble qu'elles placent la cible sous un meilleur jour⁸⁴. Donner des raisons (désir/croyance) pour expliquer une action paraît plus respectable que citer des émotions ou des traits de personnalité.

Il est sûr que ces données sont intéressantes, mais on a l'impression que parfois découvrir ce genre de politique conceptuelle est considéré comme l'aboutissement dans l'étude des théories naïves. Malle écrit comme si une reconstruction complète des politiques conceptuelles *était* la théorie naïve. Mais une théorie naïve au sens qui nous intéresse est un ensemble d'engagements théoriques qui sont présumés par le fonctionnement de la cognition, non pas la description de ce fonctionnement en elle-même.

Ce tableau a besoin d'être interprété : une image fidèle des présupposés ontologiques du sens commun ne ressort pas automatiquement. Une raison de cette difficulté est l'imbrication entre la fonction descriptive et la fonction pragmatico-stratégico-manipulative de la cognition sociale. Il est clair qu'on a besoin de décrire correctement plusieurs aspects de la réalité sociale. Si un ami veut nous trahir, si une alliance contre nous est montée ou qu'un partenaire romantique nous montre son intérêt, on a tout intérêt à le découvrir. En ce sens, la cognition sociale est soumise à des normes classiques de justification : elle vise la vérité, la description correcte d'une dimension de la réalité. En même temps, dans les interactions sociales on assiste tout le temps à une négociation de pouvoir, de rôles et de réputation. Toute affirmation, attribution, inférence ou explication nous fait avancer ou perdre des points dans ce jeu. La cognition sociale est programmée pour nous permettre en même temps de déchiffrer de façon correcte la réalité sociale et de jouer cette partie dans les dynamiques complexes du groupe. Jusque-là, rien de nouveau. D'un point de vue cognitif, la découverte intéressante est l'imbrication complète de ces deux fonctions. Il n'y a pas d'alternance entre deux états d'esprit discrets, de sorte qu'on passerait d'une attitude descriptive à une attitude de négociation et de manipulation. La fonction pragmatique est imbriquée de façon subtile dans

⁸⁴ B. F. Malle, J. M. Knobe et S. E. Nelson, "Actor-observer asymmetries in explanations of behavior: New answers to an old question," *Journal of Personality and Social Psychology* 93, no. 4 (2007).

l'usage standard des concepts de la psychologie ordinaire. Pour faire une comparaison avec l'humour, c'est comme si, au lieu d'alterner le comique et le sérieux, une ironie diffuse empêchait d'isoler clairement ce qui est sérieux et ce qui ne l'est pas.

Par exemple, dans l'attribution et l'explication par des états mentaux et de traits de caractère, le niveau de rigueur descriptive est souvent influencé aussi bien par l'accès aux connaissances pertinentes que par la motivation. Les gens ne vont expliquer « au mieux de leurs possibilités » que lorsqu'ils ont une motivation particulière à être rigoureux. Par défaut, ils vont trouver un compromis entre un niveau de véridicité acceptable et l'utilité de mettre en avant ces explications dans l'échiquier social. Même s'il y a de la vérité dans l'image d'autrui que nous construisons, souvent accepter cette image nous arrange. D'après les expériences, ces interférences pragmatiques paraissent *constitutives* de l'usage normal des concepts dans les interactions sociales ordinaires. Ce ne sont pas des exceptions : assumer une attitude complètement descriptive, détachée et rigoureuse est l'exception (et un exercice non naturel et difficile). Nous sommes programmés par défaut pour être de subtils manipulateurs sociaux. Il n'y a pas d'un côté un usage sincère et sérieux de nos concepts et de l'autre des besoins pragmatiques qui interfèrent. Les exigences de rigueur et d'efficacité conversationnelle et pratique sont emmêlées depuis le début.

Quelle est la conséquence pour la recherche des engagements ontologiques du sens commun ? On ne devrait pas croire que toutes les situations et contextes d'usage spontané, écologiquement représentatif et non *intentionnellement* manipulateur, sont sur le même niveau : dans certains cas on est programmé pour être des subtils manipulateurs, et dans d'autres non. En d'autres termes, pour arriver aux engagements ontologiques il faut ignorer ces interférences pragmatiques. Dans ce but, il faut une évaluation globale du fonctionnement du domaine cognitif.

Voici un exemple de cette dépuration : même si, comme Malle le montre, les croyances sont comparativement peu citées, on doit attribuer aux gens une psychologie populaire des croyances et désirs. Car la notion de croyance est moins mobilisée uniquement pour des raisons pragmatiques et motivationnelles. Notamment, il faut une motivation spéciale à être rigoureux. La fréquence moins importante du concept de croyance doit donc être ignorée. Ce sont les engagements théoriques épurés d'interférences pragmatiques de ce genre qui vont être pris en considération par les philosophes. Un autre exemple : souvent les explications intentionnelles sont préférées pour des raisons qui relèvent de la gestion de la réputation. Avoir des raisons explicites semble être plus respectable qu'être « poussé » par des traits de caractère. Ce genre d'avantage des explications intentionnelles doit être « décompté », car il ne touche pas aux engagements théoriques véritables. Dans un état d'esprit « sérieux » et rigoureux (créé expérimentalement en soulignant le besoin de précision), les gens ne donneraient plus cet avantage aux explications intentionnelles.

On pourrait pousser plus loin le problème posé par les interférences pragmatiques à la recherche d'engagements ontologiques. Pour arriver à comprendre la position du sens commun, il faut avoir une vision d'ensemble de la cognition sociale, interpréter le tableau qui se dessine et « décomposer » les interférences pragmatiques. Mais qu'est-ce qui se passerait si, suite à cette procédure, il ne restait rien ? C'est la position de plusieurs philosophes, qui pensent que la psychologie populaire n'a aucun engagement factuel et se résume à sa fonction pragmatique. Par exemple, Hutto et Gallagher soutiennent que les attributions de croyances et désirs sont en réalité des processus narratifs pour donner du sens au comportement de la cible, sans que cela entraîne aucun jugement factuel⁸⁵.

Les philosophes ont été plutôt sophistiqués dans la classification des différentes formes d'anti-réalisme possibles pour la psychologie populaire. Ils ont opéré une distinction claire entre une attitude réaliste (à proprement parler, le seul cas où il y a une théorie implicite), pragmatico-narrative (ex. : croyances et désirs constituent le tissu d'une narration et ne sont pas des états mentaux) et pragmatico-fictionnelle (le sens commun fonctionne « comme si » il postulait des états mentaux, mais – pour une série de raisons – un engagement ontologique sérieux semble à écarter). Le bien-fondé d'une interprétation anti-réaliste de la cognition sociale ne peut être jugé que domaine par domaine, en analysant les données dans les détails. Il faut examiner concrètement les inférences, attributions et stratégies explicatives adoptées dans un domaine donné pour établir si on peut les interpréter de façon totalement pragmatique, ou s'il y a des raisons de rejeter un engagement ontologique sérieux, favorisant ainsi une lecture fictionnaliste du sens commun (dans ce domaine). C'est ce qu'on fera dans le prochain chapitre, consacré à la théorie du caractère du sens commun.

⁸⁵ D. D. Hutto, "The Limits of Spectatorial Folk Psychology," *Mind and Language* 19, no. 5 (2004).

La théorie du caractère implicite du sens commun

Introduction et résumé

Dans ce chapitre, nous appliquerons les conclusions générales sur l'étude du sens commun du chapitre précédent à la théorie implicite de la personnalité. Pour mettre en évidence les engagements ontologiques du sens commun, il faut une cartographie globale, quantitative et qualitative, du domaine concerné. Cette cartographie doit être correctement interprétée, afin d'isoler les vrais engagements ontologiques des interférences pragmatiques. Ce sera exactement ce qu'on cherchera à faire pour la notion de trait, dans le cadre de la cognition sociale.

Après avoir montré que les emprunts très lourds au sens commun des théories philosophiques et psychologiques de la personnalité posent un problème, nous explorerons l'usage et le fonctionnement du concept de trait dans la cognition ordinaire, dans les contextes suivants : des tâches d'attribution, d'explication, de prédiction, dans les théories spontanées sur la personnalité, dans des expériences en milieu naturel et en laboratoire, ainsi qu'au fil du développement de l'enfant. Nous allons également étudier les liens entretenus avec le concept de trait par les notions de croyance, de désir et par les stéréotypes sociaux.

La conclusion de cette longue analyse sera la suivante : l'existence du concept de trait et son fonctionnement sont universels dans notre espèce. Vraisemblablement, ils font partie de notre répertoire cognitif inné. Les différences individuelles et culturelles dans l'usage des traits sont très importantes : notamment, les traits sont mobilisés plus dans les cultures occidentales « individualistes » et moins dans les cultures orientales « collectivistes ». Quelle que soit la culture, les individus qui ont une attitude individualiste les mobilisent plus et les individus qui définissent leur identité par rapport au rôle social et au groupe les mobilisent moins. Néanmoins, la meilleure explication de cette variabilité est une activation sélective : certaines cultures et quelques individus, de façon chronique et par défaut, utilisent davantage les mêmes structures cognitives universelles. L'hypothèse de l'activation sélective s'impose pour deux raisons : grâce à des procédures de *priming*, il est possible d'activer ou désactiver l'usage des traits chez tout le monde, faisant en sorte que les occidentaux commencent à les utiliser comme les orientaux, et vice-versa. Quel que soit le niveau d'activation, plusieurs contraintes dans l'usage sont respectées : cela signifie que les traits sont mobilisés plus ou moins, non pas différemment (uniquement dans une mesure limitée).

Les caractéristiques universelles du fonctionnement de la notion de trait sont les suivantes : pour peu qu'ils soient utilisés, les traits ont avant tout une fonction évaluative. Dans le processus de

négociation de la réputation sociale, il est important de décider si on peut considérer un allié, un proche ou un étranger comme étant loyaux, généraux, fiables, honnêtes, ponctuels, etc. Mais les traits ont aussi une fonction descriptive : lorsque le sens commun attribue un trait, il ne se limite pas à donner une évaluation qui pourrait être paraphrasée par « c'est quelqu'un à qui on pourra faire confiance dans nos interactions ». Il formule aussi une explication causale portant sur les processus qui produisent ce comportement fiable. Toutefois, alors que la fonction évaluative est universelle (si le concept de trait est mobilisé, il aura une fonction évaluative), les fonctions explicative et prédictive sont activées uniquement pour expliquer et prédire le comportement des inconnus. Autrement dit, les traits constituent un outil évaluatif pour tout le monde et explicativo-prédictif uniquement pour des gens peu connus. En ce qui concerne les proches et les gens qu'on connaît très bien, même si on continue à les évaluer par les traits (mon ami est sage, sympathique, gentil), en réalité on explique et prédit leur comportement grâce à toute une kyrielle de ressources et schémas conceptuels souples et contextuels.

Ce résultat montre la nécessité d'une évaluation globale d'un domaine cognitif. Si on se concentre uniquement sur la formation d'impressions sur des inconnus, les traits généraux semblent dominer la scène. En réalité, ils correspondent à une théorie explicative « locale », mobilisée uniquement pour les inconnus. Du point de vue d'une évaluation globale, la « véritable » théorie implicite du sens commun, celle qui devrait intéresser les philosophes, est beaucoup plus souple et contextuelle. Dans les termes techniques de la psychologie, on dirait « idéographique ».

Dans la deuxième section du chapitre, nous allons défendre cette conception des traits contre deux critiques. La première met en question l'interprétation réaliste et causale des traits. Les traits nous donnent une théorie explicative « locale », explicitement conçue pour expliquer le comportement des inconnus. Or, d'après certains théoriciens, les traits ne sont *jamais* explicatifs, soit parce que la psychologie populaire ne produit pas d'explications (uniquement des interprétations ou des narrations), tout court (scepticisme global), soit parce que les traits en particulier ne sont pas explicatifs (scepticisme comparatif : les notions de croyance et désir expliquent le comportement, non pas les traits). À partir des données expérimentales, nous allons montrer que les interprétations anti-réalistes de l'usage des traits par rapport aux inconnus ne tiennent pas la route.

La deuxième critique met en question l'universalité du fonctionnement des traits. Souvent, la théorie spontanée (l'ensemble des croyances explicites par rapport à un domaine) du sens commun n'est pas fiable. Par exemple, si on demande aux gens comment fonctionne la grammaire du français, on va avoir des réponses très peu fiables. Néanmoins, au fil des deux dernières décennies, Dwek a montré l'existence d'une série d'effets « top-down » de la théorie spontanée sur la pratique conceptuelle effective. Il semble que la croyance dans la stabilité (*entity theory*) ou dans la fluidité

(*incremental theory*) de la personnalité modifient profondément l'usage des concepts de traits. Notamment, ces derniers seraient mobilisés beaucoup moins par les gens qui pensent que la personnalité est fluide. Si c'était vrai, cela signifierait qu'il n'y a pas de concept substantiel et universel de trait, car la conception des traits serait profondément influencée par la culture et les convictions personnelles. À cela, nous répondons que les effets top-down existent, mais le noyau dur du fonctionnement du concept de trait reste inchangé : les traits généraux, plus au moins stables et fluides, sont des concepts essentiellement évaluatifs et gardent un rôle explicatif uniquement par rapport aux inconnus.

Dans ce chapitre, nous aurons atteint deux résultats importants. Étant donné que les théories psychologiques et philosophiques du caractère entretiennent des liens stratégiques très importants avec le sens commun, il fallait forcément expliciter les positions de ce dernier. Cela relève du sérieux, de la clarté et de la crédibilité nécessaires dans le cadre d'une science de la personnalité. Mais ce sera aussi très utile par la suite, car nous avons déjà atteint un résultat substantiel et pertinent pour une évaluation du modèle globaliste du caractère. Le sens commun a une théorie explicative globaliste (basée sur l'interaction de traits larges et très cohérents) du caractère uniquement pour les inconnus. Lorsqu'ils sont motivés et placés dans une position épistémique plus favorable, notamment avec des proches qu'ils connaissent bien, les gens emploient une théorie du caractère plus contextuelle et souple. Cela représente un gros coup pour le modèle globaliste, car c'est cette deuxième théorie qui manifeste les engagements ontologiques véritables du sens commun. L'enseignement du sens commun que le théoricien normatif du caractère devrait retenir est celui de la souplesse et complexité du caractère. Ce message n'a pas été écouté au fil de l'histoire des théories de la personnalité.

La relation problématique entre sens commun et théories de la personnalité

L'origine de la notion de personnalité dans le sens commun représente en même temps un atout et une faiblesse pour une tentative d'élaboration d'une théorie scientifique. Un atout, car la quantité et la richesse des ressources conceptuelles qu'on mobilise pour comprendre autrui sont importantes et très intéressantes à exploiter d'un point de vue théorique. Une faiblesse, car certains aspects typiques du fonctionnement des concepts communs influencent trop souvent l'approche théorique, dans le mauvais sens.

Dans un domaine donné, le sens commun est typiquement peuplé par une variété considérable de concepts et de notions qui paraissent toutes familières. Le domaine des ressources qu'on consacre à la compréhension d'autrui ne constitue certainement pas une exception. Ainsi, dans des contextes

différents, on retrouve des affirmations telles que : « il est courageux » ; « il est têtu » ; « c'est une petite bourgeoise typique » ; « c'est le profil d'un gagnant » ; « il n'a pas de caractère » ; « elle a une personnalité harmonieuse » ; « il s'est construit autour de ces valeurs » ; « ses motivations sont ce qui le caractérise le plus », etc.

Comme il arrive souvent avec les notions du sens commun, même si les relations existantes entre elles restent absolument vagues⁸⁶, de prime abord l'idée de les rejeter ou de les critiquer paraît absurde. La réaction radicale aux thèses situationnistes niant l'existence des traits de caractère est souvent l'incrédulité : « comment peut-on nier que les gens aient des traits ? Il n'y a pas de gens courageux ou timides ? ». Si la cible avait été une autre notion, parmi celles qui constituent la constellation de concepts relatifs à la personnalité, la réaction aurait été la même : « les profils de personnalité n'existent pas ? Qu'est-ce que cela signifie, qu'il n'y aurait pas des gagnants ? ». Le paradoxe du sens commun est que, malgré le fait qu'une attaque à ses notions fondamentales paraisse aller à l'encontre de l'évidence la plus élémentaire, il contient et institutionnalise par lui-même des thèses niant ses propres prémisses. Quoi de plus familier que des affirmations qui soulignent la complexité infinie et l'unicité de la personnalité de chacun d'entre nous ? Mais cela n'est pas facilement compatible avec le fait que – dans une série de contextes – nous soyons immédiatement reconnus comme étant des courageux ou des gagnants.

En somme, les différents éléments de l'apparat conceptuel que le sens commun déploie pour catégoriser et expliquer les différences individuelles entretiennent des relations très mal définies, lorsqu'ils ne paraissent pas carrément incompatibles. Malgré cela, ils restent associés à une sensation de « satisfaction explicative » qui légitime leur application dans la vie quotidienne et leur adoption dans la pratique scientifique. C'est ce sentiment de satisfaction qui fait qu'« on voit bien ce que cela signifie » lorsqu'on parle de caractère, de personnalité, de profils, de traits, de motivations et de valeurs centrales dans la structure du soi, si bien qu'on ne ressent pas le besoin d'une science de la personnalité.

Il est possible qu'on ne comprenne même pas ce que cela signifierait de travailler sur les fondements d'une science de la personnalité, tellement notre pré-compréhension du domaine paraît riche et peuplée d'instruments conceptuels efficaces. L'idée d'une étude psychologique de la personnalité évoque sans doute la nécessité d'approfondissements de questions spécifiques : est-ce que la générosité interagit de façon positive ou négative avec la timidité ? Tout le monde comprend

⁸⁶ Quel est le lien entre les traits et un profil ? la différence entre caractère et personnalité ? Comment expliquons-nous le fait qu'une personnalité soit harmonieuse, par des traits ou des motivations centrales au soi ? Construire sa propre identité autour de certaines valeurs ou motivations est-il une façon d'avoir certains traits, ou un processus compatible mais qui se place sur un autre niveau d'explication ?

la pertinence de cette question locale, mais on ne se rend pas compte qu'en réalité on fait face à une véritable crise des fondements. Avant que la science ou la philosophie puissent lui faire confiance, il faut clarifier la théorie implicite du sens commun dans les détails.

Section I - Cartographie de l'usage conceptuel : Existence et portée des concepts de trait dans le sens commun

Au cours de plusieurs décennies de recherche, la psychologie sociale et de la personnalité ont exploré le concept populaire de trait. L'hypothèse dominante – mais qui ne fait pas l'unanimité – est, comme nous l'avons dit, celle du dispositionnisme (qu'on va détailler dans le chapitre sur le situationnisme) : le sens commun cautionnerait l'existence de dispositions très générales au comportement, et ses explications du comportement d'autrui seraient largement basées sur ces dispositions. Il est inévitable que notre analyse soit centrée autour d'elle. Nous allons commencer par une revue des contextes d'usage des traits.

En ce qui concerne les données empiriques, il y a quatre types de littérature qui sont pertinents : attribution, explication, prédiction en termes de traits, et théories spontanées. Il est essentiel de garder à l'esprit que ces quatre emplois d'un concept en principe doivent être dissociés, même si en pratique il y a aura des intersections très grandes. Si je juge qu'une personne possède un certain trait de caractère (attribution), c'est probablement parce que je crois que cette hypothèse explique son comportement (explication). Cela paraît tellement évident que souvent des tâches d'attribution sont présentées directement comme des tâches d'explication⁸⁷. Mais il est clair qu'une attribution peut servir aussi des stratégies de gestion d'image, de négociation d'une réputation. Nous verrons par la suite que des tâches explicites d'explication (demander pourquoi quelqu'un a fait quelque chose) sont associées à des réponses différentes par rapport aux tâches d'attribution.

Une autre identification dangereuse est celle entre prédiction et explication : une prédiction est censée être basée sur les modèles explicatifs qu'on déploie, mais, en fait, il peut bien s'agir d'une heuristique prédictive isolée et déconnectée d'une compréhension théorique plus profonde. La divergence entre explication et prédiction dans la cognition sociale a été bien documentée par

⁸⁷ B. F. Malle. "Attributions." In *The Encyclopedia of Social Psychology*, edited by R. Baumeister and K. Vohs. Thousand Oaks, CA: Sage, 2008.

Andrews⁸⁸. Une technique de prédiction qui ne soit pas basée sur des ressources explicatives est tout simplement une forme d'induction et de reconnaissance de régularités dans le comportement. Ces régularités sont enregistrées et utilisées par le système cognitif sans être en rien expliquées. Des exemples peuvent aller d'heuristiques triviales telles que « normalement, les gens font ce qu'ils avaient dit », « les gens ont tendance à refaire ce qu'ils ont fait dans le passé », jusqu'à la reconnaissance intuitive de patterns de comportement complexes. Il n'y a pas d'accord sur la portée de prédictions que des heuristiques inductives a-théoriques de ce genre peuvent produire, mais elle pourrait être importante : il est probable que la nécessité de théorie a été exagérée. Par exemple, Goldman soutient que le potentiel inductif présent est constitué largement par des schémas et des scripts issus de l'activité de simulation passée de la vie mentale d'autrui. Dans son approche, la simulation est une activité explicative qui a des engagements ontologiques (notamment l'existence d'états mentaux tels que croyances et désirs) : cela signifie que l'induction présuppose au moins une activité d'explication passée. Autrement dit, on peut faire de l'induction aujourd'hui parce qu'on a fait de la vraie explication (via simulation) par le passé. Mais il n'y a pas de données concluantes pour étayer cette vision réductrice du rôle des heuristiques inductives⁸⁹.

Enfin, les théories spontanées devraient être fidèles au fonctionnement effectif de la cognition dans le domaine concerné, mais elles se révèlent souvent non fiables : par exemple, il suffit de penser à la non-fiabilité de la compréhension explicite et spontanée de la grammaire. Nous ne voulons pas nier que l'activité explicative est très souvent à la base d'attributions et de prédictions : seulement, les exceptions peuvent être extrêmement significatives pour l'analyse des théories implicites du sens commun.

Attribution

Juger que quelqu'un est gentil, honnête ou insolent (etc.) est peut-être l'usage le plus fréquent des termes de trait. La plupart des termes de traits sont au moins modérément évaluatifs⁹⁰, si bien que l'attribution de traits est synonyme d'« évaluation » en termes de traits. Plus que l'opposition

⁸⁸ K. Andrews. "Knowing Mental States: The Asymmetry of Psychological Prediction and Explanation." In *Consciousness: New Philosophical Perspectives*, edited by Q. Smith and A. Jokic, 201-219. Oxford: Oxford University Press, 2003.

⁸⁹ A. I. Goldman. "Interpretation Psychologized." *Mind & language* 4, no. 3 (1989): 161-185. Il faut souligner que l'opposition entre théorie de la théorie et théorie de la simulation n'est qu'une question d'architecture cognitive : dans les deux cas il y a des engagements ontologiques.

⁹⁰ D. M. Buss. "Evolutionary Personality Psychology." *Annual Review of Psychology* 42, no. 1 (1991): 459-491.

laboratoire/contexte naturel, dans cette littérature s'impose la distinction entre usage spontané et stimulé des concepts de trait, car la psychologie de la personnalité, avec près d'un siècle de questionnaires et interviews, nous a légué une quantité colossale de données sur la façon dont les gens réagissent *si on leur demande* de juger en termes de traits. Bien sûr, la visée finale de ces travaux n'était presque jamais la théorie populaire de la personnalité, mais l'étude de la personnalité tout court. Néanmoins, ces données nous montrent que les concepts de traits sont universels et ont « une vraie épaisseur cognitive ».

Dans toutes les cultures, la tâche de juger quelqu'un par des termes de traits (de sa propre langue) est familière, tout le monde comprend la question et répond sans problème. Le niveau d'accord dans le jugement de deux ou plusieurs personnes qui ne se connaissent pas forcément entre elles et qui peuvent venir de différents milieux, mais qui connaissent la cible, varie de façon cohérente avec le niveau de connaissance de la cible et l'accessibilité du trait : les gens qui connaissent très bien quelqu'un sont plus d'accord que les autres et il y a un consensus plus net sur des traits accessibles (extroversion) que sur des traits dont la manifestation est plus rare (irascibilité)⁹¹. Cela prouve que les traits ont une identité cognitive bien définie et cohérente comme forme d'organisation de l'information sur une personne : stimuler un discours sur les traits n'est pas une façon d'activer de façon générique d'autres stéréotypes ou représentations (et peut-être pas les mêmes chez des personnes et pour des traits différents, ainsi que la classe des traits serait un artefact). Nous avons affaire à une vraie espèce naturelle psychologique, une classe spécifique de représentations mentales.

Cette universalité se manifeste si l'on stimule explicitement les concepts de traits (en posant des questions explicites) : la situation de l'usage spontané est plus complexe. Il y a une grande variabilité culturelle dans l'attribution spontanée : les cultures occidentales individualistes jugent le plus en termes de traits, et la plupart des cultures dites collectivistes beaucoup moins. Si on leur demande de décrire des connaissances, les Indiens hindou font surtout référence à des rôles ou identité sociaux et des occupations, alors que les Américains multiplient l'attribution de traits généraux. La tendance à l'attribution de traits se développe progressivement de l'enfance jusqu'à l'adolescence chez les Américains mais non pas chez les Indiens⁹². Même constat pour l'auto-

⁹¹ D. T. Kenrick. "Profiting from controversy: lessons from the person-situation debate." *Personality: Critical Concepts in Psychology* 43 (1998): 23-34.

⁹² R. A. Shweder and E. J. Bourne. "Does the concept of the person vary cross-culturally?" *Culture Theory: Essays on Mind, Self and Emotion* (1984), J. G. Miller. "Cultural influences on the development of conceptual differentiation in person description." *British journal of developmental psychology* 5, no. 4 (1987): 309-319. Les données développementales sont dans Miller.

description : les Américains répondent à un test qui consiste à compléter vingt descriptions de la forme « Je suis... » (*Twenty Statement Test*) avec un grand nombre de traits, alors que les Japonais, Chinois et Coréens se concentrent plutôt sur les rôles, les appartenances et les occupations. Les étudiants américano-coréens utilisent un nombre de traits proportionnel à la proximité déclarée avec l'une ou l'autre culture⁹³.

Malgré cette variation culturelle, il semble que la fonction la plus importante des traits soit évaluative. Cela a été mis en évidence dans des études qui ont comparé l'usage quantitatif de différents concepts par rapport à des buts cognitifs différents. Les intérêts cognitifs de se souvenir de certains épisodes, éprouver de l'empathie vers une personne, l'évaluer ou expliquer son comportement ne sont pas nécessairement servis de la même façon par des concepts psychologiques naïfs tels que traits, croyances, désirs, buts ou émotions. Il pourrait y avoir une spécialisation dans la psychologie naïve.

Dans le contexte de cultures occidentales, le résultat est que les traits semblent être les concepts évaluatifs par excellence. Si on leur demande d'exprimer une opinion globale sur une personne, les gens recourent spontanément à une liste de traits, au lieu de critiquer ou louer des buts et des engagements de la cible. Dans une expérience, les participants lisaient une liste d'épisodes de comportement qui présentaient une variété intéressante et significative de références à des traits, aussi bien que des buts et situations saillantes. On leur demandait de regrouper en sous-ensembles les épisodes suivant des critères de leur choix. Le groupe dans lequel les participants s'attendaient aussi à devoir écrire une évaluation globale des protagonistes des épisodes, les catégorisaient par trait, alors que ceux qui pensaient devoir se souvenir des épisodes ou sympathiser avec le protagoniste les classaient par le genre de but impliqué⁹⁴. Il est clair qu'une visée évaluative stimule la codification de l'information en termes de traits.

Une autre expérience a répliqué le résultat que les gens mobilisent des traits lorsqu'on demande «Quel genre de personne est-il ? » et préfèrent des explications basées sur des buts pour répondre à

⁹³ Pour la confrontation avec les Coréens, E. Rhee, J. S. Uleman, H. K. Lee and R. J. Roman. "Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures." *Journal of personality and social psychology* 69, no. 1 (1995): 142-152. Pour les Chinois H. C. Triandis, C. McCusker and C. H. Hui. "Multimethod probes of individualism and collectivism." *Journal of Personality and Social Psychology* 59, no. 5 (1990): 1006-1020, G. W. M. IP and M. H. Bond. "Culture, Values, and the Spontaneous Self-Concept." *Asian Journal of Psychology* 1, no. 1 (1995): 30-36. Pour les Japonais, S. D. Cousins. "Culture and self-perception in Japan and the United States." *Journal of Personality and Social Psychology* 56, no. 1 (1989): 124-131.

⁹⁴ C. Hoffman, W. Mischel and K. Mazze. "The Role of Purpose in the Organization of Information About Behavior - Trait-Based Versus Goal-Based Categories in Person Cognition." *Journal of Personality and Social Psychology* 40, no. 2 (1981): 211-225.

la question « Pourquoi vous pensez qu'il a pu agir de cette façon ? »⁹⁵. Comme nous le verrons par la suite, les cultures collectivistes utilisent en général moins les traits, mais en gardant les mêmes proportions entre différents contextes d'usage : il n'arrive jamais que les traits soient centraux pour l'explication mais non pas pour l'attribution. Ils peuvent être globalement moins employés, mais la hiérarchie entre les différents types d'usage est respectée. Cela confirme l'universalité du rôle cognitif des traits (pour peu qu'ils soient activés).

Prédiction

Si l'attribution/évaluation est la première fonction des concepts de traits dans la psychologie naïve, la prédiction occupe certainement la deuxième place. En fait, nous serions tentés de donner la première place au rôle que jouent les traits dans la prédiction, car nous retrouvons moins de variabilité inter-culturelle que pour l'attribution.

Le paradigme expérimental le plus utilisé est une tâche de prédiction du comportement futur pertinent pour un trait de caractère, à partir de l'exposition au comportement passé sous la forme de brèves histoires, vidéos ou rencontres réelles. On demande aux participants quelle est la probabilité que la cible se comporte d'une certaine façon dans le futur. Souvent, la description du comportement passé et futur n'est pas fournie en termes de traits, pour ne pas influencer le résultat. Aussi bien les Occidentaux que les Asiatiques semblent coder massivement le comportement passé en termes de traits et ils formulent des prédictions de façon conséquente. Par exemple, à partir d'une histoire où il se passe beaucoup de choses et il y a aussi une action courageuse (mais pas présentée en tant que telle) du protagoniste, les gens considèrent comme très probable pour le futur un cours d'action qui contient des actions courageuses (mais pas présentées en tant que telles). Cela prouve qu'au niveau cognitif le comportement est codé partout en termes de traits.

La différence intéressante est que les Asiatiques sont en mesure de moduler les prédictions basées sur les traits à partir de l'information situationnelle saillante : quelqu'un d'extroverti par le passé sera extroverti aussi dans une nouvelle situation, mais davantage si la situation s'y prête particulièrement (une fête) et moins dans une situation inappropriée (dans une église). Les Asiatiques prennent en considération cette information situationnelle alors que les Occidentaux projettent de façon mécanique le niveau passé de manifestation du trait dans le futur⁹⁶. Mais, au

⁹⁵ K. Lemm and Y. Shoda. "Dispositional inference type varies with consistency of observed behavior and judgment task." Unpublished work (1995).

⁹⁶ A. Norenzayan, I. Choi and R. E. Nisbett. "Cultural Similarities and Differences in Social Inference: Evidence from Behavioral Predictions and Lay Theories of Behavior." 28, no. 1 (2002): 109-120.

bout du compte, les Asiatiques eux aussi activent des traits dans la gestion de la tâche de prédiction. Cela prouve l'universalité du rôle prédictif des traits. Pourquoi alors ne pas affirmer que la prédiction est la fonction principale des traits ?

Le problème est que le rôle prédictif dans le cas de personnes familières, importantes et qu'on connaît bien est beaucoup plus douteux. Alors qu'il est clair que les traits gardent leur fonction évaluative aussi bien pour les étrangers que pour les proches (le consensus dans l'évaluation de nos proches dans des questionnaires le prouve), des données directes sur le rôle prédictif avec les proches sont difficiles à trouver. En général, les commentateurs n'ont pas considéré une dissymétrie possible par rapport à la familiarité, peut-être à cause du lien par défaut entre attribution, prédiction et explication. Etant donné que les traits sont mobilisés pour des tâches d'attribution avec des proches, la prédiction et l'explication ne pouvaient que suivre.

À cause de la conviction que les traits sont massivement utilisés pour la prédiction, et étant donné que, dans le comportement effectif des gens, il y a moins de cohérence (inter-situationnelle) que l'existence des traits ne le prédit, il y a eu des débats complexes sur la façon dont les traits pouvaient aider (ou donner l'impression d'aider, ou juste cacher leur inadéquation radicale) pour prédire le comportement d'amis et de proches (voir chapitre sur le situationnisme : étiologie de l'erreur). Si les traits que le sens commun invoque n'existent pas, comment les gens peuvent-ils les employer dans la vie de tous les jours pour anticiper les actions de leurs proches?

Plusieurs types de biais d'exposition et de confirmation sélectives ont été postulés, aussi bien que la possibilité d'une influence directe dans l'attitude de nos proches. Je pense qu'un ami est timide, modeste et ponctuel : si ce genre d'information était presque tout ce que j'utilise, je devrais générer des prédictions inadéquates, mais, en fait, je ne suis pas déçu parce que je vois mon ami avant tout dans des situations qui stimulent la timidité (exposition biaisée), j'ignore des faits qui ne supportent pas l'hypothèse (biais de confirmation), ou je produis avec ma présence une attitude timide chez mon ami (prophétie auto-réalisatrice)⁹⁷. L'existence de ces effets est beaucoup plus certaine que leur capacité douteuse de cacher complètement les incohérences dans une image de l'ami qui serait construite essentiellement à l'aide de traits.

L'attitude consistant à donner pour acquis le rôle des traits dans la prédiction du comportement de nos proches est un signe de la domination de la littérature sur la formation d'impressions en psychologie de la personnalité et en psychologie sociale. La façon dont les gens réagissent à la première rencontre, entretien, ou épisode raconté est censée révéler les éléments de base de la

⁹⁷ J. M. Doris. *Lack of character: personality and moral behavior*. Cambridge, U.K. ; New York: Cambridge University Press, 2002, K. Andrews. "It's in your nature: a pluralistic folk psychology." *Synthese*: 1-17.

cognition sociale. Ensuite, on s'intéresse très peu à montrer comment tout cela peut produire la texture riche d'une compréhension personnelle profonde. Cette attitude est injustifiée : la distinction « connaissance superficielle/connaissance profonde et attachement émotionnel » a fait ses preuves comme « espèce naturelle psychologique ». Il y a des processus et structures cognitives consacrés aux relations et aux personnes significatives. Que ceux qu'on connaît et qu'on aime fassent une différence au niveau cognitif a été soupçonné depuis des siècles et a été pleinement confirmé dans plusieurs littératures contemporaines, par exemple les études sur le style d'attachement⁹⁸. Il est donc au moins bizarre que l'idée dominante du primat prédictif des traits soit basée entièrement sur la littérature de formation d'impressions, ignorant le facteur familiarité.

Il y a sans doute beaucoup de simplification et de vision stéréotypée dans la compréhension de nos proches (ex. : le compagnon qui est l'amant parfait ou le fils qui est l'étudiant modèle), et cela est évident dans littérature abondante sur les relations. Mais cette vision stéréotypée ne peut être que locale. Pour s'en convaincre, il suffit de rappeler les hypothèses extrêmes que les défenseurs du rôle prédictif des traits ont besoin d'invoquer. Dans leur perspective, il faut que ce soit vrai que nous voyons nos amis dans un ensemble extrêmement limité de situations ; autrement il serait effectivement incroyable que nous soyons à ce point aveugles à leur complexité comportementale⁹⁹. La littérature sur les schémas relationnels en cognition sociale a montré que nous avons un dossier cognitif foisonnant consacré à chacun de nos proches, contenant une grande variété de structures conceptuelles¹⁰⁰. Nous associons tout un paquet d'idéaux, buts, situations et sentiments typiques (que la personne concernée a ou que nous avons lorsqu'on est avec elle) et aussi traits à chaque relation. Alors que les traits typiques d'un ami sont sans doute activés dans le cadre de ce schéma relationnel (ex. : penser au meilleur ami active les trait d'honnêteté, loyauté, gentillesse), dans une visée évaluative, il est difficile d'évaluer leur rôle prédictif. On sait beaucoup de choses sur le comportement des personnes importantes pour nous : on a des connaissances directes, situationnelles, détaillées et même des statistiques intuitives sur ce qu'ils ont fait, leurs motivations, etc. Il est probablement vrai que certaines prédictions sont basées sur des traits, surtout dans les parties simplifiées et stéréotypées de l'image que nous avons de nos proches, mais il n'y a pas de données empiriques pour soutenir que la contribution est significative. Les données sur le rôle dans

⁹⁸ M. D. S. Ainsworth. *Patterns of Attachment: A Psychological Study of the Strange Situation*: Lawrence Erlbaum Associates, 1978.

⁹⁹ Doris. *Lack of character: personality and moral behavior*.

¹⁰⁰ M. W. Baldwin. "Relational schemas and the processing of social information." *Psychological bulletin* 112, no. 3 (1992): 461-484.

l'explication qu'on va passer en revue maintenant vont étayer davantage le scepticisme quant au rôle des traits dans la compréhension des personnes importantes et bien connues.

Explication

La notion d'explication est rendue opérationnelle en psychologie par des réponses à des questions où on demande « pourquoi » les gens font ce qu'ils font. Les traits enregistrent leur pire résultat dans des contextes explicatifs. D'abord, la fréquence absolue d'explications qui mobilisent des traits est très basse par rapport à ce qu'on imagine. Une méta-analyse de 8 études sur des explications spontanées montre qu'elles n'apparaissent que 5% du temps¹⁰¹. Nous avons vu que si on modifie une question qui demande une évaluation d'une personne en une question explicative, on a moins de réponses qui mobilisent des traits¹⁰². Le but d'évaluer, mais non pas celui d'expliquer, améliore la mémorisation d'épisodes qui avaient été présentés de façon neutre¹⁰³.

Des études qui avaient trouvé un usage beaucoup plus important des traits dans les explications dans les cultures occidentales, jusque 40%, en réalité se référaient à tous les concepts dispositionnels confondus et opposés à une référence à des facteurs dans la situation¹⁰⁴. Donc le 40% comprend aussi croyances et désirs : la spécificité de la notion de trait a été ignorée dans la plupart de ces études, et lorsque cela a été contrôlé de façon rigoureuse, comme dans l'étude de Malle, leur part paraît mince. Les psychologues n'ont pas remarqué cette asymétrie entre attribution et explication car, comme nous l'avons vu, ils ont souvent présumé que les attributions sont des explications : si cela était vrai, alors on explique tout le temps par des traits, mais cela devient un mystère de comprendre pourquoi toute référence disparaît lorsqu'on demande explicitement « pourquoi » les gens font ce qu'ils font.

Le deuxième problème est l'idée selon laquelle la seule espèce naturelle psychologique qui se manifeste dans les tâches d'explication serait la distinction entre d'un côté la situation et de l'autre tout ce qui est « à l'intérieur du sujet ». Cette approche a guidé la recherche pendant des décennies,

¹⁰¹ B. F. Malle, J. M. Knobe and S. E. Nelson. "Actor-observer asymmetries in explanations of behavior: New answers to an old question." *J Pers Soc Psychol* 93, no. 4 (2007): 491-514.

¹⁰² Lemm and Shoda. "Dispositional inference type varies with consistency of observed behavior and judgment task." Unpublished.

¹⁰³ Hoffman, Mischel and Mazze. "The Role of Purpose in the Organization of Information About Behavior - Trait-Based Versus Goal-Based Categories in Person Cognition."

¹⁰⁴ J. G. Miller. "Culture and the Development of Everyday Social Explanation." *Journal of Personality and Social Psychology* 46, no. 5 (1984): 961-978.

en mettant sur le même plan et dans le même décompte quantitatif des concepts tels que croyances, désirs, intentions et traits. Nous commenterons davantage l'inexistence d'une distinction naïve entre personne et situation dans le chapitre sur le situationnisme.

Les problèmes ne se terminent pas ici, car il faut prendre également en considération certains effets pragmatiques. Il semble qu'être dans une position épistémique meilleure (meilleure connaissance de la personne) et motivé à être rigoureux n'augmente pas l'usage de traits, et pourrait même le réduire. Dans les expériences de Malle, les traits ne sont pas mobilisés davantage avec des proches. Dans une expérience de Mischel, ils le sont légèrement moins¹⁰⁵, tout cela *sans* exclure (notamment, dans les expériences de Malle) les explications nuancées et contextuelles telles que « timide en ce qui concerne... » ou « courageux dans le contexte de... ». Dans certaines études, des attributions de ce genre semblent représenter jusqu'à 20% du total¹⁰⁶. Elles devraient être éliminées car « être timide dans une certaine situation » n'est pas équivalent à être timide tout court, à posséder le trait de la timidité.

Mischel pense que parfois même des occurrences qui n'ont pas une forme linguistique conditionnelle pourraient en réalité se référer à des traits nuancés et contextuels, par des mécanismes pragmatiques (ex. : cela est implicite dans le contexte que « timide » ici signifie « timide avec le chef ») ou parce que des connaissances explicatives plus riches sont pour ainsi dire « résumées » dans un label de trait : par exemple, des étudiants qui avaient un profil très idiosyncrasique mais stable du caractère consciencieux (*conscientiousness*) (ex. : très consciencieux en classe, dans les résidences, mais pas dans la cour de récréation ou pendant des examens) se considéraient très consciencieux, alors que d'autres étudiants qui étaient en moyenne aussi consciencieux que les premiers mais avec un profil individuel instable (pas toujours consciencieux ou pas dans le même type de situation) ne se considéraient pas consciencieux. Dans ce cas, il semble que le mot « consciencieux » se réfère à une connaissance très riche sur soi-même, non pas à un trait général¹⁰⁷.

La tendance à expliquer moins par les traits le comportement des personnes qu'on connaît mieux n'a pas été toujours répliquée : par exemple, Hapson a trouvé une tendance opposée dans une tâche

¹⁰⁵ L. C. Idson and W. Mischel. "The personality of familiar and significant people: the lay perceiver as a social-cognitive theorist." *Journal of Personality and Social Psychology* 80, no. 4 (2001): 585-596.

¹⁰⁶ J. C. Wright and W. Mischel. "Conditional hedges and the intuitive psychology of traits." *Journal of personality and social psychology* 55, no. 3 (1988): 454-469.

¹⁰⁷ W. Mischel and P. K. Peake. "Beyond Deja-Vu in the Search for Cross-Situational Consistency." *Psychological Review* 89, no. 6 (1982): 730-755.

de description libre de personnes significatives ou de simples connaissances¹⁰⁸. Le problème est qu'il s'agit de descriptions qui sont avant tout une forme d'évaluation, attribution, jugement. La description est une tâche évaluative un peu hybride suivant la manière dont on l'interprète, pour cela il est important de contrôler et tester explicitement des explications, non pas des descriptions. En ce sens, à part les expériences de Malle, le travail de Chen est remarquable. Dans trois expériences, Chen a testé comment les gens expliquent des comportements spécifiques de la forme « si...(ex : il est dans une fête), alors ...(ex : il s'isole dans un coin) » pour des proches, des connaissances ou des stéréotypes sociaux (l'avocat, le scientifique, etc.). L'idée de base est de formuler une liste d'épisodes de cette forme et ensuite les expliquer comme on veut. Les variations de l'expérience étaient : compléter avec un « alors » la situation de départ déjà donnée (ex : si un vieux copain l'appelle, alors....) ou formuler entièrement les épisodes comportementaux (ex : si..., alors...); dans le premier cas, compléter obligatoirement le « alors » ou seulement si on est confiant sur la bonne réponse (ex : on sait que notre meilleur ami s'isole dans un coin dans les fêtes); viser des situations générales (dans une fête...) ou spécifique (si on lui demande ce qu'il fait dans sa vie, alors...); en ce qui concerne les explications, les formuler obligatoirement ou seulement si on est confiant, et mesurer le niveau de confiance.

D'abord, ces expériences montrent une différence claire entre contenu descriptif et explicatif dans la cognition sociale. La qualité de la connaissance explicative ne suit pas automatiquement la richesse de la connaissance descriptive. Il y a des cas (environ 30%) où on connaît très bien un comportement spécifique en ignorant l'explication possible (omise ou marquée avec un bas niveau de confiance). Ensuite, les explications par les traits diminuent de façon nette pour les personnes significatives (dont, par ailleurs, on connaît plus d'épisodes et on est plus certain de leur explication) par rapport aux stéréotypes et connaissances. Malheureusement, l'étude – après avoir exclu les pseudo-explications (des affirmations faites avec le ton d'une explication mais qui en fait ne font que paraphraser la régularité comportementale de départ ou expliquer comment on la connaît : « parce que je l'ai vu le faire hier ») – opère une distinction uniquement entre les explications qui mobilisent des états mentaux proximaux (émotions, buts, croyances, désirs) et les autres (traits et situations). La part des « autres » tombe de 66% (connaissances) à 41% (proches). Dans les explications « autres », il ne semble pas plausible que le pourcentage des traits augmente pour les proches, l'auteur l'aurait signalé comme un effet remarquable. Il paraît donc qu'on

¹⁰⁸ S. E. Hampson. "Trait ascription and depth of acquaintance: the preference for traits in personality descriptions and its relation to target familiarity." *Journal of research in personality(Print)* 17, no. 4 (1983): 398-411.

explique moins par les traits le comportement des proches, alors qu'on devrait s'attendre à l'effet inverse¹⁰⁹.

Traits et stéréotypes

Une autre raison pour douter du rôle explicatif des traits vient de la littérature sur les stéréotypes. On a longtemps pensé que les traits jouaient un rôle « fondationnel » dans la cognition sociale, au sens où ils seraient à la base de la typologie intuitive : l'ensemble des différentes façons de classer les gens en types, sur des bases sociales ou autre. Prenons le stéréotype du scientifique : on l'imagine avec des lunettes et une veste blanche, mais – on a longtemps pensé – le noyau dur des prédictions et des explications qui dérivent de la mobilisation du stéréotype seraient le résultat du fait qu'un scientifique est considéré (disons) comme intelligent, consciencieux et ordonné. Si cette approche était correcte, il faudrait créditer les traits de toute la puissance explicative et prédictive des stéréotypes. La thèse est empiriquement testable, car elle fait des prédictions sur la structure des stéréotypes existants et possibles : on ne devrait pas retrouver des thèmes décisifs et qui ne peuvent pas dériver du jeu combinatoire entre les différents niveaux des traits.

Or, il semble que dans la cognition sociale les traits ne jouent pas ce rôle fondateur. On a cru longtemps que c'était le cas, sur la base de l'association claire et fiable entre un stéréotype et un certain nombre de traits. Il y a un accord considérable sur la manière de répartir une liste déterminée de traits dans un nombre donné de types humains, et le même stéréotype évoque de façon fiable la même liste de traits chez tout le monde. Mais cela n'est pas suffisant en soi, car il est hors de doute que les stéréotypes sont associés à des traits : un scientifique est diligent, studieux, déterminé. Il faut évaluer le poids descriptif et explicatif des traits dans la structure du stéréotype.

En ce qui concerne le contenu descriptif, il est désormais clair que les stéréotypes sont beaucoup plus riches qu'un ensemble de traits. On associe régulièrement à un stéréotype donné des attentes, des émotions, des attitudes, des caractéristiques physiques, des situations typiques (où on le rencontre, où il se retrouve souvent). Par exemple, les personnes âgées sont associées à l'idée de répondre lentement et de marcher le dos courbé, d'autres classes sociales passent leur temps assis derrière un bureau de travail ou parlent beaucoup¹¹⁰. Tout de même, cette richesse de détail n'est

¹⁰⁹ S. Chen. "Psychological-State Theories about Significant Others: Implications for the Content and Structure of Significant-Other Representations." *Personality and Social Psychology Bulletin* 29, no. 10 (2003): 1285.

¹¹⁰ J. A. Bargh, M. Chen and L. Burrows. "Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action." *Journal of Personality and Social Psychology* 71, no. 2 (1996): 230-244, S. M.

pas vraiment surprenante : il n'y a pas de raisons pour lesquelles les gens ne devraient pas remarquer directement une série de détails dans le comportement moyen dans une profession, classe sociale ou tranche d'âge et l'associer à leur stéréotype. Après tout, les personnes âgées répondent souvent plus lentement que les jeunes. Au-delà d'une couche de détails factuels qui peuvent être ajoutés directement, l'important est de vérifier que, au niveau explicatif et prédictif, le noyau dur qui tient ensemble le stéréotype est une liste de traits. Une façon de tester cela est de voir si les effets de l'activation des stéréotypes dans l'activité cognitive sont dus à l'activation des traits correspondants. Avons-nous une conception riche et spécifique de la façon dont les personnes âgées sont lentes, qui auraient des effets spécifiques sur le traitement de l'information dès qu'elle est activée, ou faut-il considérer que le stéréotype « vieux » active uniquement l'idée de lenteur qui aura ses effets génériques ? Les psychologues ont longtemps cru à la deuxième hypothèse : « *Lorsqu'on rencontre une personne âgée, on active la catégorie de « vieux » et des traits associés tels que « lent ». Dans tous les cas, l'activation de la catégorie « lent » aurait des effets sur le comportement, indépendamment de comment elle a été activée* »¹¹¹.

Mais récemment de nouvelles données sont venues soutenir la première. Dans une tâche de reconnaissance de mots ou non-mots (décider si une série de signes est un mot existant) faite après une tâche de reconnaissance de visage de personnes âgées (à partir d'une série de photos de gens de tout âge présentée sur l'écran), on s'attendrait à ce que l'activation du stéréotype de « vieux » fasse répondre les participants plus lentement (que le groupe de contrôle). En revanche, si le mot à reconnaître était lié au trait de « lent », la vitesse de réponse aurait dû augmenter (pas nécessairement jusqu'au niveau du groupe de contrôle).

Les attentes ont été confirmées, mais la chose intéressante est que, dans une série de variations, les effets sur la vitesse de réponse du stéréotype « vieux » et du trait « lent » lui-même étaient significativement différents, même s'ils allaient dans le même sens (temps de réaction plus longs). En d'autres termes, en ce qui concerne la lenteur, l'influence de l'activation du stéréotype de « vieux » n'est pas identique à celle de l'activation de « lent » : on a une représentation spécifique de la façon dont ils sont lents qui est cognitivement active et ne dérive pas automatiquement de l'association du trait « lent » au stéréotype de « vieux »¹¹².

Andersen and R. L. Klatzky. "Traits and social stereotypes: levels of categorization in person perception." *Journal of personality and social psychology* 53, no. 2 (1987): 235-246.

¹¹¹ A. Dijksterhuis, H. Aarts, J. A. Bargh and A. van Knippenberg. "On the Relation between Associative Strength and Automatic Behavior." *Journal of Experimental Social Psychology* 36, no. 5 (2000): 531-544.

¹¹² K. Kawakami, H. Young and J. F. Dovidio. "Automatic Stereotyping: Category, Trait, and Behavioral Activations." *Personality and Social Psychology Bulletin* 28, no. 1 (2002): 3.

Non seulement les traits n'expliquent pas les conséquences de l'activation cognitive des stéréotypes, mais le rapport pourrait être inversé. Il y a des données qui montrent que les traits pourraient être des généralisations à partir du contenu plus riche de stéréotypes. Au lieu de partir d'une notion déjà donnée de courage qui aurait plusieurs manifestations possibles, il y aurait d'abord une série d'attitudes spécifiques face au danger chez des figures stéréotypées : la façon remarquable dont un soldat, un médecin ou un martyr affrontent le danger physique. Le trait du courage serait une généralisation à partir de ces attitudes, au sens où chaque manifestation du courage serait empruntée à l'un des stéréotypes. Lorsque je juge qu'une action est courageuse, je mobiliserais l'un de ces contenus plus riches : je trouve qu'elle est courageuse au sens d'un soldat, d'un médecin, etc¹¹³.

Globalement, il ne semble pas que les traits jouent un rôle explicatif majeur : l'occurrence moyenne est faible, et cela ne s'améliore pas lorsqu'il devrait, c'est à dire lorsqu'on est « sérieux » d'un point de vue explicatif (bonne connaissance et motivation à être rigoureux). Si on expliquait le comportement d'autrui par une théorie postulant l'interaction complexe de traits (ex. : il a fait cela parce qu'il est courageux, honnête et patient), cela devrait au moins se voir avec nos proches et dans la génération de stéréotypes. Or, il n'y a pas de signes de cela.

Scepticisme quant au rôle des concepts de trait dans la psychologie populaire

Les traits sont une famille de termes qui existent dans chaque langue connue et sont typiquement utilisés pour qualifier des personnes et des actions. La cognition sociale liée aux traits a été considérée comme importante par une grande majorité de traditions théoriques en philosophie (la théorie de la vertu dans toutes ses formes depuis Confucius et Aristote) et en psychologie (psychologie de la personnalité traditionnelle et une majorité des approches contemporaines).

¹¹³ Andersen and Klatzky. "Traits and social stereotypes: levels of categorization in person perception." Dans l'expérience, on fait évaluer sur une échelle de 0 à 10 le niveau d'association de tout couple dans l'ensemble formé par deux traits (introverti, extroverti), six stéréotypes (comédien, homme politique, gourou etc.) et 31 attributs (bien habillé, cultivé etc). Par exemple, comédien-cultivé 7, homme politique-bien habillé 8. Le résultat est un réseau d'associations qui a la propriété suivante : pour un trait donné, les attributs qui lui sont le plus associés font aussi partie de l'ensemble des attributs qui sont le plus associés à un certain stéréotype, alors que l'inverse n'est pas vrai. Il y a des stéréotypes pour lesquels l'ensemble des attributs qui leur sont le plus associés n'est partagé par aucun trait. D'après les auteurs, cela signifie que tout le contenu des traits vient probablement des stéréotypes, qui sont plus riches.

Comme il arrive souvent, à des positions majoritaires anciennes et très enracinées s'opposent des manifestations de scepticisme minoritaires, mais également anciennes et combatives.

Il y a deux formes de scepticisme quant aux traits. Un scepticisme *tout court* (ou absolu) : l'idée selon laquelle les explications par les traits seraient des pseudo-explications. En disant « il l'a fait parce qu'il est courageux », on a bien l'intention d'expliquer des actions, mais en réalité on serait en train de ne rien expliquer. Nous ne discuterons pas davantage ici cette forme de scepticisme, car il s'agirait d'évaluer la théorie naïve des traits au lieu de la décrire, ce qui n'est pas le but du présent chapitre. Une forme de scepticisme plus pertinent est l'éliminativisme ou scepticisme *comparatif* quant aux traits. Le discours sur les traits constituerait uniquement une façon de parler pour se référer à d'autres phénomènes psychologiques, typiquement des schémas complexes de croyances et désirs (un trait est un ensemble de croyances et désirs, d'après Brandt)¹¹⁴. Nous allons analyser cette forme de scepticisme à partir des données empiriques.

Il y a beaucoup de données qui montrent que mobiliser les termes de trait n'est pas une « façon de parler » superficielle à laquelle ne correspond aucune réalité cognitive intéressante. Ces données ont été recueillies à partir de tâches et méthodes différentes, dans des milieux naturels, dans le laboratoire et au fil du développement de l'enfant. Le discours sur les traits ne peut être réduit ou paraphrasé facilement par d'autres concepts naïfs tels que croyance, désir ou intention : il y a une véritable pensée sur les traits dans le sens commun. En particulier, après avoir passé en revue les données, nous pourrions exclure les hypothèses suivantes : (1) la cognition relative aux traits est identique à la psychologie populaire de croyances et désirs (réduction) ; (2) elle constitue une façon plus complexe et coûteuse de parler des mêmes phénomènes (hypothèse flatteuse), ou alors (3) uniquement une façon pratique mais imprécise (sub-optimale) de les aborder (hypothèse banalisante). D'après (2), les concepts de traits sont difficiles à maîtriser, ils requièrent une grande maturité cognitive, mais en dernière instance se réfèrent uniquement à des interactions complexes de croyances et désirs. Même si le réseau de croyances et désirs qui constitue le courage est difficile à saisir (il est difficile de raisonner sur le mode d'action de la générosité ou de l'honnêteté), aucune structure cognitive nouvelle ne serait postulée par le discours sur les traits. (3) représente l'hypothèse contraire. Au lieu d'un concept qui résume de façon raffinée des interactions trop complexes à expliciter, les traits nous donneraient une façon simple d'articuler de manière approximative des explications en termes de croyances et désirs. Dans ce cas de figure, pour être rigoureux, il faudrait citer directement les croyances et les désirs correspondants : les traits constitueraient uniquement un raccourci facile mais approximatif.

¹¹⁴ R. B. Brandt. "Traits of Character: A Conceptual Analysis." *American Philosophical Quarterly* 7 (1970): 23-37.

La réduction entre des concepts universels et importants du sens commun est très improbable

Un premier commentaire non décisif est que l'hypothèse de la réduction complète entre des notions centrales du sens commun n'a plus guère de crédibilité¹¹⁵. Les concepts analysables sont périphériques et spécialisés : par exemple, le concept de célibataire pourrait être éliminé, en le remplaçant par la négation de marié (non marié(e))¹¹⁶ : plus on se dirige vers des notions centrales et massivement utilisées et moindres paraissent les chances d'une réduction. Or, nous avons montré que les traits sont des structures d'organisation de connaissances universelles associées à des pratiques inférentielles et attributives spécifiques : ils remplissent donc les conditions dans lesquelles la possibilité de réduction ne paraît pas crédible.

Dissociation entre traits et croyances/désirs dans l'ontogenèse

La différence dans l'ontogenèse entre l'apparition de la capacité d'attribuer des croyances et des désirs et celle d'attribuer des traits est un autre argument contre l'identification des deux activités cognitives. Entre la troisième et la quatrième année, les enfants deviennent pleinement capables d'attribuer des croyances et des désirs, mais ils ne maîtrisent pas encore l'usage des concepts de trait pour prédire le comportement futur. Ils peuvent comprendre et reconnaître des actions liées à un trait (ils savent ce que ces mots signifient et peuvent reconnaître des actions gentilles, malhonnêtes ou loyales) mais semblent incapables d'utiliser le comportement passé pour prédire l'attitude future de façon conséquente : par exemple, ils ne s'attendent pas à ce qu'on soit gentil parce qu'on a agi de façon gentille par le passé). Cette capacité se développe au moins trois ans plus tard (dans le milieu naturel : voir ci-dessous pour des expériences de « choix forcé » où la compétence est plus précoce)¹¹⁷.

¹¹⁵ T. Williamson. *Knowledge and Its Limits*: Oxford University Press, 2000.

¹¹⁶ La réduction est problématique même dans ce cas : par exemple, nous ne dirions pas qu'un bébé est célibataire.

¹¹⁷ W. S. Rholes and D. N. Ruble. "Children's Impressions of Other Persons: The Effects of Temporal Separation of Behavioral Information." *Child Development* 57, no. 4 (1986): 872-878, N. Yuill and A. Pearson. "The Development of Bases for Trait Attribution: Children." *Developmental Psychology* 34, no. 3 (1998): 574-586. Pour une présentation générale W. S. Rholes, L. S. Newman and D. N. Ruble. "Understanding self and other: Developmental and motivational aspects of perceiving persons in terms of invariant dispositions." In *Handbook of Motivation and Cognition: Foundations of Social Behavior*, edited by E. T. Higgins and R. M. Sorrentino, 369-407. New York: Guilford Press., 1996.

Dissociation entre traits et croyances/désirs

Il y a au moins un cas net de dissociation entre l'usage de traits et croyances/désirs. Les enfants autistes, qui ne maîtrisent pas l'attribution de croyances et désirs, peuvent apprendre à reconnaître des actions pertinentes pour un trait et faire des prédictions conséquentes grâce à la *Social Story Therapy*¹¹⁸. Rien ne semble capable de leur faire maîtriser l'usage de croyances et désirs de la même façon.

Si le décalage dans l'ontogenèse exclut l'hypothèse de la réduction, la dissociation exclut l'hypothèse (flatteuse) selon laquelle les traits seraient uniquement une façon plus complexe et exigeante (d'un point de vue cognitif) pour raisonner sur d'autres états mentaux.

Les explications par des traits ne sont pas non plus des moyens simplifiés pour parler de croyances et désirs. Cette forme de scepticisme quant aux traits ne cherche pas à éliminer ce genre de concept par une réduction ou par l'idée que ce serait une façon plus complexe et générale de parler d'états mentaux. Au contraire, il s'agirait d'un raccourci pratique, simplificateur et peu rigoureux pour parler de choses plus complexes. Si cela était le cas dans l'économie de la cognition sociale, la prédiction serait que les gens devraient abandonner la référence à des traits lorsqu'ils deviennent « sérieux » d'un point de vue explicatif (ils ont la connaissance adéquate et la motivation épistémique). Cette hypothèse a été testée. Dans la même série de huit expériences qu'on a déjà citée, Malle, Knobe et Nelson ont comparé systématiquement (entre autres) la quantité d'explications de traits, d'autres facteurs personnels et dispositionnels (émotions, buts) et de croyances et désirs dans des explications spontanées écrites ou enregistrées. La distribution de ces explications a été étudiée dans une variété impressionnante de conditions différentes : explication d'épisodes choisis (par le participant) et dont on se souvient bien, qui concernent soi-même ou d'autres, d'épisodes choisis par l'expérimentateur, l'explication à des inconnus ou à des proches d'épisodes qui concernent soit des proches soit des inconnus (lesquels peuvent ou pas avoir été avec nous au moment de l'épisode concerné), le tout avec ou sans la motivation de représenter la cible sous un beau jour¹¹⁹. Presque toutes les variables qui peuvent intervenir dans une explication ont été manipulées : épisode expliqué, personnes impliqués dans l'épisode, personnes à qui on explique, motivation à expliquer de façon rigoureuse.

Dans cette analyse exceptionnelle, le seul facteur qui semble réduire l'usage des traits en faveur de croyances et désirs était la motivation à donner une représentation positive de la personne

¹¹⁸ Andrews. "It's in your nature: a pluralistic folk psychology.", C. Gray and F. Horizons. *Writing Social Stories with Carol Gray*: Future Horizons, 2000.

¹¹⁹ Malle, Knobe and Nelson. "Actor-observer asymmetries in explanations of behavior: New answers to an old question."

impliquée. Comme nous l'avons vu, il semble que les explications intentionnelles sont plus flatteuses, elles affichent un côté plus flatteur de la vie mentale d'un individu. Cela n'est pas difficile à comprendre, car des raisons intentionnelles qui prennent la forme de croyances et désirs sont normalement associées à la subjectivité et rationalité.

En tout cas, les gens semblent considérer l'action des traits comme une partie spécifique de l'activité mentale qui n'est pas au même niveau que les états mentaux intentionnels. Les données ne soutiennent pas l'attitude sceptique suivante : « bon, tu m'as dit qu'il est courageux, soyons sérieux et explique-moi pourquoi il l'a fait ». Si c'était le cas, la motivation à être rigoureux devrait faire baisser le nombre d'explications par les traits (qui d'ailleurs, nous le répétons, sont très rares). Cet effet a été confirmé pour d'autres cas, mais pas pour les traits. Par exemple, les mêmes études prouvent que dans les explications intentionnelles la mobilisation du concept de désir est facile et fréquente, alors qu'il faut davantage de connaissance de la cible et de la motivation épistémique pour reconstruire correctement les états de croyance. Cela montre que les explications intentionnelles basées presque entièrement sur l'attribution de désirs sont une version simplifiée des vraies explications que les sujets savent produire, lorsqu'ils sont sérieux d'un point de vue explicatif. Vraisemblablement, les explications par les traits et les explications intentionnelles n'ont pas ce genre de rapport entre elles : les explications par les traits ne sont pas aux explications par les croyances et désirs ce que les explications par les désirs (uniquement) sont aux explications par les croyances et désirs.

Statut

Une dernière forme de scepticisme quant au rôle des traits dans le sens commun concerne leur statut. Par rapport aux traits, la psychologie du sens commun pourrait être considérée comme étant réaliste (causalité mentale : les traits sont des dynamiques réellement actives dans l'esprit des gens), interprétativiste-narrative ou fictionnelle. D'après une approche narrative, la psychologie populaire produit des narrations qui remplissent une série de fonctions liées à l'efficacité communicative et pragmatique, sans postuler l'existence réelle des entités et des processus cités dans les narrations. « Sa volonté et sa détermination l'ont poussé à ce choix » serait une narration qui ne postule pas l'existence causale de la volonté et de la détermination. L'approche fictionnelle reconnaît que les engagements théoriques de la psychologie populaire *semblent* sérieux : tout se passe « comme si »

les gens postulaient vraiment l'existence d'états mentaux ; ou presque, car certaines attitudes des gens trahissent l'absence d'engagement ontologique véritable¹²⁰.

Je ne discuterai pas des positions non-réalistes qui résultent d'une simple application aux traits de scepticismes plus généraux quant à l'explication psychologique, indépendamment des données empiriques. Par exemple, je ne discuterai pas une forme de « béhaviorisme analytique » général (l'idée selon laquelle tous les concepts psychologiques se limiteraient à énoncer l'existence de dispositions au comportement, sans les expliquer). Je crois que la détermination du statut d'un concept naïf dans la cognition sociale doit être basée sur des données spécifiques relatives à l'usage et au rôle cognitif qu'il joue. Néanmoins, dans le camp même des approches empiriques, il y a une distinction entre des positions hétérodoxes (non réalistes) globales quant à la psychologie populaire et des positions hétérodoxes uniquement quant aux traits.

Anti-réalisme global

Commençons par les anti-réalistes globaux : lorsqu'on regarde les données, sans forcer des engagements ontologiques indépendants, cette position n'est pas très convaincante. Qu'il s'agisse d'une forme d'interprétativisme (narratif ou non) ou de fictionnalisme, les approches hétérodoxes globales cherchent à montrer que les gens ne sont pas sérieusement engagés d'un point de vue ontologique et explicatif par rapport aux concepts naïfs de la cognition sociale. L'argument consiste à repérer des signes que les gens parlent de façon métaphorique ou sans réagir de façon correcte aux données pertinentes (*responsiveness to evidence*), ce qui devrait être le cas s'ils étaient sérieux d'un point de vue explicatif.

Or, les cas qui sont le fond de commerce des interprétativistes et fictionnalistes existent certainement. Souvent, les gens ne modifient pas une explication psychologique pour chercher davantage d'adéquation factuelle, mais simplement pour augmenter l'impact narratif de leurs propos. Egalement, on trouve souvent des emplois clairement métaphoriques ou fictionnels pour améliorer l'efficacité communicative : « il croit que les femmes sont toujours disponibles pour écouter » ou « il veut le soutien de tous les étudiants » peuvent être des attributions utilisées pour résumer une attitude, une situation, sans véritables engagements ontologiques sérieux. En d'autres termes, on ne croit pas que *toutes* les femmes sont disponibles à écouter, et on ne cherche pas

¹²⁰ Typiquement, on entretient une attitude fictionnaliste par rapport à des thèses qu'il serait très coûteux de rejeter : par exemple, le libre arbitre. Tout se passe comme si on croyait vraiment dans le libre arbitre. Voir S. Smilansky. *Free Will and Illusion*: Clarendon Press, 2000.

véritablement le soutien de *tous* les étudiants. On croit peut-être que beaucoup de femmes dans un ensemble raisonnablement large de situations sont disposées à écouter, mais dire « toutes » accomplit une série de fonctions liées à l'efficacité pragmatico-communicationnelle.

L'existence de ces emplois non littéraux ne prouve pas grand chose : d'abord, la précision de toute représentation de la réalité dans la vie quotidienne est soumise à des contraintes d'efficacité narrative et pragmatique. Si je raconte un dîner romantique à un ami, je peux faire court et dire qu'il n'y avait que deux chaises dans la cuisine, alors qu'en réalité il en avait trois mais la troisième reste toujours inutilisée à côté du frigo et ne constitue pas un vrai élément de la scène. Que le sens commun s'autorise cette liberté communicative ne met pas en question son réalisme sur les chaises ou d'autres objets physiques. En ce qui concerne le fictionnalisme, tout concept peut être utilisé *aussi* de façon fictionnelle ou métaphorique pour résumer un état de chose ou une attitude (ex. : mon vieux portable est un dinosaure). Deuxièmement, il y a des contextes clairs, stables et larges dans lesquels toutes les contraintes d'une attitude réaliste sont remplies. À propos d'une dispute, d'un événement traumatique passé ou dans un tribunal les gens discutent très sérieusement sur ce qu'une telle personne croyait, désirait, voulait ou savait à un moment donné. Des doutes sur ce qu'un ami perdu avait ou non compris de nous peuvent nous tourmenter pour des décennies. Dans toutes ces situations, les gens paraissent aussi sérieux « ontologiquement » que possible.

Peut-être la coexistence « pacifique » de contextes narratifs, métaphoriques et réalistes est plus problématique que ce que nous laissons entendre. Godfrey-Smith montre qu'il y a beaucoup de contextes dans la vie quotidienne dans lesquels les gens sont beaucoup moins engagés ontologiquement que dans un tribunal, et cela paraît un phénomène intéressant d'un point de vue cognitif. Autrement dit, la proportion des contextes métaphoriques est si élevée que la psychologie populaire semble conçue *aussi* pour fonctionner de cette façon. Godfrey-Smith réagit à cette situation en devenant pluraliste sur le statut de la psychologie naïve : parfois les gens sont sérieux ontologiquement, parfois ils sont fictionnalistes. Cela correspond à l'usage des modèles en science, et Godfrey-Smith propose de considérer la psychologie naïve comme une forme de modélisation. Le point important est qu'un modèle soit heuristiquement fructueux : une interprétation réaliste de sa structure est optionnelle, elle peut être adaptée à des contextes d'usage et non pas à d'autres. « *Le modèle de la psychologie populaire n'impose pas son interprétation. Si on demande « Quel est l'engagement de la psychologie populaire en elle-même ? », la réponse est « aucun*¹²¹ » ».

¹²¹ P. Godfrey-Smith. "Folk psychology as a model." *Philosophers' Imprint* 5, no. 6 (2005): 1-16.

Nous croyons que le pluralisme de Godfrey-Smith est basé sur une confusion. Il est vrai que l'usage métaphorique est constitutif de l'usage correct : il est programmé en profondeur et prévu par dessein dans la cognition sociale. Il ne relève pas simplement de la possibilité contingente d'utiliser de façon narrative ou métaphorique n'importe quel concept. Si le pluralisme signifie seulement cela, alors nous sommes d'accord. Mais nous n'avons pas encore touché à la question de l'engagement ontologique. Imaginons que dans une culture donnée, parler d'amour soit, selon les situations, une façon d'exprimer un engagement authentique ou une manière de manipuler pour obtenir des faveurs. Supposons aussi que cette double face soit elle-même essentielle pour les dynamiques qui maintiennent l'équilibre social dans cette culture. Or, ce serait difficile d'argumenter que cette culture n'est pas engagée ontologiquement dans l'existence de l'amour authentique. De la même façon, si les gens sont réalistes pour des choses sérieuses et importantes (une dispute, dans un tribunal), du point de vue des engagements ontologiques ils sont réalistes tout court.

Anti-réalisme spécifique aux traits

Il y a aussi un scepticisme comparatif quant au statut des traits par rapport à d'autres concepts de la cognition sociale. Il y aurait un engagement ontologique à propos des croyances, désirs, intentions, mais pas par rapport aux traits. Celle-ci est la position que nous défendons, mais nous voulons discuter une façon non satisfaisante de la soutenir : une analyse hétérodoxe et « non mentaliste » du concept de trait. Analyse et engagement ontologique par rapport à un concept ne sont pas la même chose. « Sorcière » peut bien signifier ce que tout le monde imagine (analyse) et néanmoins être utilisé par un sujet qui n'y croit pas et n'est pas sérieux sur l'existence des sorcières (engagement ontologique). Cela est à mon avis le cas pour les traits : il est correct d'analyser le concept naïf de trait comme une certaine dynamique mentale causalement active, mais le système cognitif social humain n'est pas ontologiquement engagé dans l'existence véritable de traits.

Une autre voie possible pour arriver au même résultat de désengagement ontologique est de proposer une analyse hétérodoxe behavioriste du concept naïf : un trait, par exemple « courageux », ne se réfère qu'à des régularités dans le comportement ; aucun phénomène psychologique qui produit ces régularités ne ferait l'objet de ce qui est « dit » lorsqu'on parle de traits. Courageux est juste quelqu'un qui affiche un pattern approprié d'actions courageuses. Ce pattern doit être expliqué

et n'est pas explicatif en lui-même¹²². Nous verrons que ce genre d'interprétation behavioriste des traits se retrouve aussi hors du contexte de la théorie du caractère naïve, dans la discussion sur la théorie psychologique des traits, car des théoriciens soutiennent qu'on devrait utiliser l'interprétation behavioriste et c'est cette notion qui aurait un travail intéressant à faire en psychologie¹²³.

Dans les deux cas, nous ne sommes pas convaincus. Andrews soutient que les traits sont seulement des heuristiques prédictives et non pas des concepts mentaux comme croyances et désirs. Nous serions contraints de lui donner raison, si nous ne connaissions pas les données portant sur la distinction entre inconnus et proches. En fait, de premier abord l'usage ontologiquement désengagé paraît très répandu : une affirmation telle que « il est courageux » souvent peut être interprétée comme un simple élément narratif, une évaluation positive avec des références factuelles très floues, un commentaire pour rendre la situation plus intelligible (identifier cette personne au stéréotype de courageux peut aider à visualiser l'enjeu de la situation de manière plus efficace) ou juste une recommandation à s'attendre probablement à certaines actions de la part de cette personne (identifier cette personne au stéréotype de courageux et cautionner les prédictions stéréotypées qui en découlent sera plus efficace que ne rien assumer).

La prolifération des usages désengagés semble faire en sorte que le « sauvetage réaliste » appliqué aux croyances et désirs ne soit pas possible, car il n'y aurait pas de contexte d'usage ontologiquement engagé clair et stable comme le tribunal. Mais à bien y regarder, je crois que ce contexte existe : il s'agit notamment de l'usage des traits avec des inconnus. L'usage métaphorique et fictionnel est largement confiné aux proches et personnes bien connues. Dans ce cas, à partir de données que nous avons discutées, on s'attend déjà à ce que les traits ne jouent pas un rôle important dans la prédiction et explication, mais plutôt dans l'évaluation et la description. Pour les inconnus, il me semble que les contraintes requises pour une « phénoménologie réaliste » soient suffisamment remplies. La littérature sur la formation des impressions montre que les gens ont des attentes très fortes de cohérence dans un sens pertinent avec le trait qu'ils ont attribué ; ils font des prédictions conséquentes et sont très surpris lorsque les attentes sont démenties par les faits. Ce sont des signes suffisants pour conclure qu'à ce moment-là leur système cognitif entretient l'attribution du trait de façon sérieuse et engagée.

¹²² On a dit qu'on ne discuterait pas le behaviorisme analytique général, mais il s'agit ici d'une forme de behaviorisme analytique spécifique aux traits.

¹²³ D. M. Buss and K. H. Craik. "The act frequency approach to personality." *Psychological Review* 90, no. 2 (1983): 105–126.

En tout cas, la différence entre des traits behavioristes et des traits causaux-explicatifs a été étudiée dans la littérature empirique. Les conclusions sont les suivantes : (1) les deux types de concepts de traits coexistent ; (2) une majorité des traits sont causaux-explicatifs mais une minorité dont l'usage est presque entièrement behavioriste est bien identifiable ; (3) les traits causaux-explicatifs se développent progressivement à partir de traits uniquement behavioristes, qui commencent à être mobilisés par des enfants de 4-5 ans. Nous avons vu que les enfants commencent à utiliser les traits de façon compétente très tard : même à l'âge de 11 ans, on ne retrouve pas encore toutes les régularités typiques de l'usage adulte. En particulier, les cinq facteurs du *big five* qu'on reconstruit à partir des questionnaires ne paraissent pas encore avoir la même structure largement confirmée chez les adultes¹²⁴. La maturation est lente. Mais quelles en sont les phases ? Jusqu'à il y a une vingtaine d'années, la question n'avait pas été analysée dans les détails et on envisageait une augmentation progressive de la maîtrise dans toutes les applications des concepts de trait. On s'est après rendu compte qu'il y a une cristallisation progressive, précoce mais plutôt rapide d'une version behavioriste des traits, à laquelle succède une maturation plus lente qui produit une compréhension causale-explicative des mêmes.

Le paradigme expérimental pertinent opère une confrontation entre la capacité prédictive (à partir du comportement passé qui est pertinent pour un trait) du comportement futur et de réactions émotionnelles futures (face à un événement) qui ne soient pas associées à des actions précises (par exemple, être déçu ou embarrassé). Par exemple, quelqu'un qui a volé une pomme (passé) va voler un vélo (comportement futur), ou éprouver du remord (réaction émotionnelle future). Un trait behavioriste n'est pas concerné par des réactions émotionnelles qui ne sont pas liées évidemment à l'action, alors que la conception causale-explicative ordinaire des traits donne une place à des simples réactions émotionnelles, qui seront à leur tour par la suite liées au comportement.

Une compétence sélective dans la prédiction du comportement, mais non des émotions, prouve la possession d'un trait behavioriste peu concerné par des processus psychologiques internes. Néanmoins, il faut bien montrer que la compétence dans la prédiction du comportement est basée sur une conception authentique du trait (behavioriste) : si les enfants arrivaient à faire des prédictions correctes avec d'autres heuristiques, le décalage développemental entre traits behavioristes et causaux serait effacé. En particulier, il est possible que les enfants se basent sur une simple évaluation de similitude entre les deux situations : si quelqu'un a volé une pomme par le passé, il pourra voler un vélo, il s'agit du même type d'action. Si c'était le cas, les enfants devraient

¹²⁴ E. M. Donahue. "Do children use the Big Five, too? Content and structural form in personality description." *Journal of Personality* 62, no. 1 (1994).

avoir une mauvaise prestation dans la prédiction d'un comportement couvert par le trait, mais éloigné par rapport à celui du passé : ils ne devrait pas forcément prévoir que quelqu'un qui a volé une pomme va tricher à un examen (honnêteté). Il est aussi possible que les enfants font des prévisions uniquement basées sur la valence : qui a fait de bonnes/mauvaises actions par le passé va faire le même type d'action dans le futur.

Dans une expérience remarquable, Yuill a fait lire à des enfants de groupes d'âges différents (4, 5, 6 et 7 ans, plus un groupe d'adultes de contrôle) 16 histoires présentant globalement 8 couples de traits opposés (ex. : honnête/malhonnête). Pour chaque épisode, relatif au comportement pertinent pour un des 16 traits, les enfants devaient répondre à 5 questions qui vérifiaient : (1) l'identification du trait (ex : reconnaissance que le vol relève de la malhonnêteté) ; (2) inférence au comportement similaire (voler pomme/voler vélo) ; (3) inférence au comportement éloigné (voler pomme/tricher à un examen) ; (4) inférence à des réactions émotionnelles face à un événement futur (être choisis pour le rôle principal dans une pièce de théâtre à l'école va générer de la joie/peur dans un enfant extroverti/introverti) et (5) inférence au comportement pertinent pour un trait différent mais de la même valence (malhonnêteté, égoïsme)¹²⁵.

Ce paradigme expérimental permet d'exclure un succès dans la prédiction de l'enfant qui utiliserait des règles aussi simples que ce qui est désirable socialement, ce qu'il ferait lui-même (grâce à la présence de traits opposés qui devraient générer la même prédiction si on suit ces règles) ainsi que la similarité (question sur le comportement éloigné) et la valence (si on utilise cette règle, on devrait prévoir que le malhonnête est forcément égoïste). Le résultat montre qu'à partir de performances mauvaises à quatre ans, on arrive très rapidement à une bonne compétence dans la compréhension behavioriste « authentique » (l'usage systématique de règles plus simples est exclu) des traits, alors que la prédiction des réactions émotionnelles, même si elle dépasse le niveau du hasard très rapidement, mature de façon plus lente et progressive¹²⁶. En plus, il semble que les traits où la partie « interne » émotionnelle est plus importante sont maîtrisés plus tard¹²⁷.

En tout cas, un usage purement behavioriste (où par courageux on voudrait uniquement dire « il défie souvent le danger ») des traits même à l'âge adulte est certainement possible, par exemple on pourrait juste appeler courageux les étudiants qui ont choisi des sujets d'examens plus difficiles.

¹²⁵ Yuill and Pearson. "The Development of Bases for Trait Attribution: Children." *Expérience* 1.

¹²⁶ Il faut rappeler qu'il s'agit d'expériences de laboratoire à choix forcé : l'usage spontané des traits dans des contextes écologiquement valides à 5-6 ans reste très réduit.

¹²⁷ N. Yuill. "Children's production and comprehension of trait terms." *British journal of developmental psychology* 10 (1992): 131-142.

Mais cela est trivial, car presque tout concept peut être utilisé de façon métaphorique ou désengagée.

Plus intéressant, une minorité de traits sont effectivement behavioristes en eux-mêmes et à tout âge : par exemple, bavard. Dans des expériences où les participants attribuent des traits à partir de plusieurs épisodes racontés sur un personnage fictif, il a été montré que l'évaluation des traits behavioristes est sensible uniquement à la fréquence des actions pertinentes (pour bavard, combien il a parlé, toute circonstance confondue), alors que pour les traits causaux à parité de fréquence, l'évaluation change, signe qu'il y a une évaluation de la situation mentale interne (ex : il a fait un cadeau, mais peut-être pas parce qu'il est généreux)¹²⁸. Exception faite pour ce groupe limité de traits behavioristes, les traits dans l'usage adulte ont une valeur causale-explicative, au sens où ils sont un genre de concept qui *peut* être utilisé dans une explication. Ce n'est pas une erreur conceptuelle de le faire, comme ce serait le cas s'ils étaient tous behavioristes à la façon de « bavard ». Cela, comme nous l'avons montré, ne signifie pas qu'ils sont utilisés systématiquement dans les explications : on peut avoir dans sa boîte un outil qu'on utilise peu, et spécialement dans des circonstances particulières (en ce cas, l'explication du comportement des inconnus).

Théorie spontanée des traits

Nous avons montré ci-dessus que l'un des sens légitimes de théorie implicite est celui de théorie spontanée, c'est-à-dire un ensemble de représentations réellement encodées dans le système cognitif, qui portent sur le phénomène en question. Si l'existence de ces théories est un phénomène cognitif intéressant en soi-même, l'erreur possible est de considérer ces théories comme étant représentatives des vrais engagements ontologiques.

Normalement, cette erreur n'est pas fréquente : nous n'avons guère la tentation de prendre au sérieux les reconstructions explicites de la grammaire ou la croyance d'un « raciste dans les faits » qui ne se considère pas raciste. Mais en ce qui concerne les traits, l'erreur représente un risque plus important, car il y a de bonnes raisons pour tomber dans le piège.

Il y a deux raisons pour lesquelles une théorie spontanée pourrait être prise au sérieux. Première possibilité : vu que ce qui est décisif est l'usage concret d'un concept, elle pourrait représenter l'usage de façon fidèle, comme si elle était le résultat d'une espèce d'introspection et analyse

¹²⁸ L. K. Kammrath, R. Mendoza-Denton and W. Mischel. "Incorporating if... then... personality signatures in person perception: Beyond the person-situation dichotomy." *Journal of Personality and Social Psychology* 88, no. 4 (2005): 605–618.

psychologique spontanée de bonne qualité sur la pratique conceptuelle. La deuxième possibilité est que la théorie spontanée influence et modifie l'usage, par un effet top-down.

La première possibilité n'est pas préoccupante : l'auto-compréhension de l'usage d'un concept ne s'est jamais révélée un processus véritablement fiable et il est hors de question de mettre en doute une reconstruction empiriquement détaillée des engagements ontologiques parce qu'elle ne colle pas avec une théorie spontanée. À différence de la grammaire d'une langue, la cognition sociale n'est pas neutre d'un point de vue motivationnel : une théorie spontanée donnée peut être soupçonnée d'être là pour des raisons pragmatiques. Avoir une certaine vision de la personnalité peut produire des conséquences pratiques dans l'interaction sociale. Ce constat vient s'ajouter aux doutes généraux sur la fiabilité de l'auto-analyse.

Par contre, une littérature empirique récente illustre la possibilité d'effets top-down en ce qui concerne la perception de la personnalité. En général, les effets top-down sont le danger numéro un pour un programme de recherche sur les engagements ontologiques du sens commun. Si les engagements ontologiques exprimés dans la pratique conceptuelle sont issus de l'influence de croyances explicites (au lieu de l'inverse), on peut encore parler d'engagements théoriques du système cognitif du sujet, mais ce genre d'engagement a des chances d'être moins stable, répandu et intéressant d'un point de vue philosophique. Moins stable parce que si la théorie explicite change, la politique conceptuelle peut suivre ; moins répandu car il peut y avoir une forte variation culturelle (et même à une échelle infra-culturelle) dans les croyances explicites, si celles-ci ne sont pas stabilisées par des « universaux cognitifs » ; moins intéressant car, comme nous l'avons vu, du point de vue d'une approche populiste du sens commun, les engagements non stables ne sont pas d'une grande utilité normative¹²⁹. Si on devait retrouver des effets top-down très forts un peu partout dans la cognition sociale, l'idée même de découvrir des théories naïves, telle qu'elle est conçue en psychologie et philosophie, n'aurait plus de sens.

La puissance des théories spontanées du caractère est ce que le programme de recherche de Carol Dweck cherche à prouver. Dweck a introduit la distinction entre la théorie de l'entité (*entity theory*) et la théorie incrémentale (*incremental theory*) de la personnalité. D'après la théorie de l'entité, la personnalité serait stable, fixe et difficile à modifier. Inversement, pour la théorie incrémentale, la personnalité est plus flexible et variable. Etant donné qu'il s'agit de théories spontanées, il faut tester ce que les gens pensent explicitement. Ainsi, le simple questionnaire introduit par Dweck est composé par une série d'affirmations de la théorie de l'entité, par exemple : « *Les personnes*

¹²⁹ Le philosophe n'est pas dans la position de force de « revendre » au sens commun ses positions modifiées et améliorées, si les gens entre-temps on changé leur idée de départ.

peuvent faire en sorte que les autres les apprécient, mais elles ne peuvent pas vraiment changer leur personnalité ». Les participants doivent signaler leur niveau d'adhésion. On évite de partir d'affirmations de la théorie incrémentale, car parler de la possibilité de changement pour tout le monde est trop « politiquement correct » et invite une adhésion généralisée¹³⁰.

L'importance de l'adhésion à des théories spontanées de ce genre aurait paru douteuse si ce n'était que Dweck et ses collaborateurs ont consacré les deux dernières décennies à prouver de façon rigoureuse l'existence d'une liste impressionnante de corrélats :

La théorie incrémentale est associée à la capacité à exploiter des informations saillantes sur la situation pour corriger l'attribution des traits à des inconnus. Lorsque le fait que la situation force un choix est saillant, les gens cessent d'attribuer l'attitude correspondante au sujet (voir ci-dessus). La plupart des gens dans les cultures collectivistes et une minorité (les allocentriques) dans les cultures individualistes sont en mesure de le faire, et ils semblent avoir une théorie incrémentale de la personnalité. Potentiellement, la variation dans la théorie spontanée pourrait donc expliquer entièrement la variation culturelle qui a été documentée dans l'usage des traits.

La théorie de l'entité est associée à un usage maximal des traits (et à davantage de confiance dans l'attribution, à parité de choix attributifs) dans les tâches d'attribution et prédiction. Par exemple, les « théoriciens de l'entité » à partir d'un épisode concernant une action « courageuse » passée sont plus certains que la moyenne dans : l'attribution du courage, la prédiction d'actions courageuses futures, la conviction que le trait est stable dans le temps et a une grande cohérence inter-situationnelle (les mesures de la confiance étant une estimation de probabilité d'une action future ou une auto-évaluation explicite du niveau de confiance)¹³¹.

La théorie incrémentale par rapport à soi-même est bénéfique dans le processus d'auto-régulation et permet, entre autres : de s'inspirer de ceux qui ont fait mieux plutôt que de se consoler avec ceux qui ont fait pire face à un échec ; de réduire l'effet de l'intériorisation de stéréotypes sociaux négatifs et faux (*stereotype threat*) ; de chercher à dépasser sa propre timidité au lieu de la rationaliser ; de s'engager pour dépasser des obstacles et problèmes dans une relation durable, au

¹³⁰ C. S. Dweck, Y. Hong and C. Chiu. "Implicit Theories Individual Differences in the Likelihood and Meaning of Dispositional Inference." *Personality and Social Psychology Bulletin* 19, no. 5 (1993): 644.

¹³¹ Attribution, prediction et coherence inter-situationnelle, voir C. Y. Chiu, Y. Y. Hong and C. S. Dweck. "Lay dispositionism and implicit theories of personality." *Journal of Personality and Social Psychology* 73, no. 1 (1997): 19-30. Attribution, G. D. Heyman and C. S. Dweck. "Children's thinking about traits: Implications for judgements of the self and others." *Child Development* 69, no. 2 (1998): 391-403. Stabilité temporelle, C. A. Erdley and C. S. Dweck. "Children's Implicit Personality Theories as Predictors of Their Social Judgments." *Child Development* 64, no. 3 (1993): 863-878.

lieu de rentrer dans une spirale de récriminations. La théorie incrémentale a un effet protecteur par rapport à la tristesse et permet d'utiliser les sentiments négatifs comme une incitation à progresser. Globalement, cet ensemble très vaste d'effets est cohérent. L'idée selon laquelle la personnalité peut évoluer a trois types de conséquences psychologiques : une motivation générale à s'améliorer de façon continue, une tendance à concevoir des problèmes et des expériences négatives comme une opportunité pour évoluer¹³² et, ce qui nous intéresse le plus dans ce chapitre, une modération quantitative et qualitative de l'usage des traits et d'autres stéréotypes qui interviennent dans la formation des impressions.

Les deux premiers effets généraux découlent de façon plutôt évidente de la croyance dans la flexibilité de la personnalité, mais quel est le lien avec l'usage limité des traits ? C'est une question d'investissement cognitif : si la personnalité ne change pas, essayer de découvrir ces facettes est un investissement excellent, car la compréhension produite sera toujours utile dans l'interaction sociale avec la personne concernée. Ainsi, la théorie de l'entité est associée à une concentration sur les traits et maximise leur mobilisation, dans toutes les tâches pertinentes. Au contraire, si la personnalité évolue, découvrir toutes les facettes à un moment donné est un investissement beaucoup moins intéressant : les traits pourraient changer très rapidement et l'effort de compréhension aura été inutile.

Comment réagir à la puissance de ces effets top-down ? Comme nous l'avons dit, le risque de mettre à mal le projet d'une étude de la théorie implicite du sens commun est important. Le fait que tous ces effets soient corrélés de façon nette à la théorie implicite du sens commun identifiée par le questionnaire de Dweck est un argument très puissant en faveur d'une direction de causalité qui va de la théorie spontanée à la pratique conceptuelle. Mais alors, pourquoi ne serait-il pas possible de modifier la pratique inférentielle à son dessein, à partir de la diffusion de nouvelles « idéologies » explicites de la personnalité ? Si tout ce qui est en jeu sont des idéologies de la personnalité qui se stabilisent plus ou moins bien suivant le milieu culturel ou social, la philosophie n'aurait pas grand chose à apprendre du sens commun. Pas même le philosophe qui n'a pas d'ambition universaliste et se concentre uniquement sur sa propre culture : car, si tout ce qui tient ensemble la pratique conceptuelle d'une culture par rapport à la personnalité est une idéologie, rien n'assure que les facteurs de stabilisation de ce discours sont destinés à perdurer. Des idéologies qui n'ont pas d'ancrage profond dans la cognition humaine peuvent changer ou disparaître très rapidement.

¹³² L'existence d'un problème rend un projet d'auto-amélioration plus saillant et facile à planifier. On n'est pas seulement moins motivé à s'améliorer si tout va plus ou moins bien, mais c'est aussi plus difficile de décortiquer la situation et d'isoler les points sur lesquels il convient de travailler plus.

Heureusement, je ne crois pas que la vulnérabilité complète de la cognition sociale (relativement à la personnalité) à des effets top-down soit la bonne interprétation des données. En fait, l'influence ne concerne que la dimension stabilité-flexibilité des traits et tout ce qui en découle, mais tout le reste paraît relever de dynamiques cognitives universelles. Même dans le cas des cultures occidentales où la théorie de l'entité est dominante, les traits paraissent des concepts essentiellement évaluatifs, centraux dans la négociation de statut et réputation sociales, avec un rôle prédictif important pour les inconnus et un rôle explicatif très limité. Cette hiérarchie d'usages n'est subvertie nulle part, on ne trouve pas, par exemple, une culture où le concept de trait serait central dans l'explication des proches, où le rôle explicatif serait crucial et le rôle évaluatif marginal. Par rapport aux cultures occidentales, qui présentent la mobilisation maximale du concept de trait, la variation existante ailleurs dans le monde relève tout simplement d'une mobilisation inférieure : sous l'influence de thèmes culturels relatifs à la fixité ou stabilité de la personnalité, des cultures différentes ont un recours plus ou moins important à une notion de trait qui garde le même profil et structure partout (à part la question de la fixité). Avec plus ou moins de confiance, si l'information situationnelle n'est pas suffisamment saillante, les mêmes erreurs de prédiction et attribution ressortent partout.

Les influences culturelles top-down ne réécrivent pas à leur gré le fonctionnement de la cognition sociale : en ce qui concerne les rapports possibles entre nature et culture, on est plutôt dans le cas de figure de structures innées universelles susceptibles d'être plus ou moins activées par des inputs environnementaux et culturels. Des données récentes le prouvent. Il paraît que l'adoption d'une approche fixiste ou incrémentale peut varier d'un trait à l'autre (on peut être fixiste sur la timidité et incrémental sur la gentillesse) et que l'attitude générale n'a pas une grande stabilité temporelle : à distance de plusieurs semaines, si on répète le test qui détecte la théorie spontanée, on a le même résultat moins souvent qu'on ne le croyait¹³³. Cela signifie que les théories de l'entité et incrémentale correspondent à des attitudes cognitives présentes chez tout le monde et que les différences culturelles et individuelles qui ont été documentées doivent être interprétées par l'idée d'activation chronique.

On a activation chronique d'une structure cognitive par rapport à des structures concurrentes lorsque la première correspond à la réaction par défaut, plus fréquente, facile et accessible dans le système, les autres étant toutefois possibles et disponibles. L'influence culturelle rend chronique l'attitude fixiste ou incrémentale dans un individu, mais l'autre approche reste facilement activable

¹³³ C. S. K. Poon and D. J. Koehler. "Lay personality knowledge and dispositionist thinking: A knowledge-activation framework." *Journal of Experimental Social Psychology* 42, no. 2 (2006): 177-191.

et son développement n'est pas empêché par le climat culturel hostile. On est clairement dans une conception de la cognition sociale comme « boîte à outils innés surabondants » que la culture se limite à mobiliser de façon sélective. Il est vrai que le statut chronique d'une des deux attitudes est lui-même très solide, mais on doit néanmoins rejeter (dans ce cas) le modèle d'une culture qui serait capable d'élaguer la panoplie d'outils innés potentiellement disponibles, de sorte que, au bout du développement, certains manqueraient à l'appel.

En tout cas, la variation culturelle et les effets top-down documentés ne touchent pas à la thèse centrale de ce chapitre : il n'y a pas de théorie de traits dans l'engagement ontologique du sens commun. Aussi stables ou flexibles soient-ils, des traits généraux n'ont pas de véritable rôle explicatif, dans aucune culture.

La possibilité d'une science de la personnalité

Introduction et résumé

La variation psychologique individuelle désignée par la notion commune de personnalité est un phénomène complexe et stratifié. La définition du cadre conceptuel d'une science de la personnalité qui fasse autorité et qui ne se limite pas à décrire une liste de faits, ou à représenter « le point de vue des psychologues », rencontre plusieurs obstacles. L'existence d'un tel cadre théorique est une condition nécessaire pour faire démarrer la collaboration interdisciplinaire entre psychologie et théorie de la vertu.

Pour que la psychologie de la personnalité soit prise au sérieux par les philosophes, il ne suffit pas de résoudre le problème épineux du lien entre être et devoir-être (premier chapitre), ni de clarifier les liens problématiques que les notions théoriques de trait ou de type entretiennent avec le sens commun (deuxième et troisième chapitres). La conceptualisation de la personnalité pose un véritable problème théorique pour la psychologie, pour deux raisons :

Une tendance à la simplification et au réductionnisme : chaque tradition ou théoricien se concentre sur un niveau d'analyse de la personnalité, cherchant à tout expliquer à partir de là. Cette démarche réductionniste oppose deux approches : une tradition qui souligne l'efficacité et la cohérence des manifestations du caractère, conçu sous le modèle de la compétence pratique (*skill*) et une autre qui identifie le caractère à la fixation des valeurs d'une personne et d'une image de soi. Dans le premier cadre, d'inspiration aristotélicienne, la définition de l'identité ne peut que suivre la compétence pratique : une personne qui devient déterminée dans le fait finira par se reconnaître en tant que telle. L'essentiel est de se concentrer sur la cohérence et l'efficacité de l'action. Dans le deuxième, ce qui compte, c'est la fixation des valeurs : chez quelqu'un qui a intériorisé un idéal de détermination, des actions qui manifestent systématiquement la détermination peuvent suivre ou non, selon les pressions de la situation ou la faiblesse humaine. L'essentiel est le processus d'individuation et de fixation des valeurs.

L'ambition d'unifier l'étude de la personnalité et du soi : la plupart des psychologues de la personnalité ont tendance à intégrer les représentations et les dynamiques cognitives qui définissent l'individualité ou le « soi » - ce qui rend chaque personne unique - dans une définition très large de personnalité. Dans ce contexte, le critère de succès d'une science de la « personne » serait d'expliquer *en même temps* la personnalité et le « soi ». Il s'agit d'un critère trop ambitieux. Si on le retient, il faut admettre que le cadre conceptuel d'une science de la personne n'est pas encore établi.

Ce chapitre sera consacré à une proposition de solution à ces deux problèmes.

En ce qui concerne le premier, il faut reconnaître que le foisonnement d'approches théoriques différentes et incompatibles qui caractérise l'histoire de la réflexion sur la personnalité tient à la complexité et au caractère pluridimensionnel du phénomène étudié. En ce sens, face à une réalité complexe et stratifiée, tant qu'on n'a pas trouvé un cadre conceptuel adéquat, il est impossible d'éviter le réductionnisme. Sans le soutien d'un cadre conceptuel fort, les théories qui démultiplient les entités et les processus paraissent *ad hoc* et sans véritable pouvoir explicatif. Ainsi, parmi les entités et les processus dont le rôle dans le façonnement de la personnalité est plausible (tempérament, traits, types, valeurs, buts, projets, défis personnels, émotions, sentiments, croyances, désirs, etc.), à chaque fois les théories de la personnalité ont fait un choix, cherchant à maximiser le rôle d'un sous-ensemble et à minimiser le rôle joué par le reste. Cette situation a contribué à la perte de crédibilité de la psychologie de la personnalité : des théories *centrées* sur les traits se sont opposées à des théories centrées sur les valeurs, ou les buts, ou les projets de vie censés définir et stabiliser la personnalité de l'individu. Aucune approche ne semble avoir les moyens de l'emporter et il est difficile de ne pas voir de l'arbitraire dans ces débats.

Le cadre conceptuel nécessaire pour éviter le réductionnisme nous est offert par la théorie de l'évolution. L'hypothèse d'une architecture cognitive « stratifiée » typiquement évolutionniste s'accorde très bien avec nos connaissances empiriques sur le fonctionnement de la personnalité : comme nous le montrerons dans les détails dans la deuxième et troisième partie, aucune tradition « réductionniste » (que ce soit tout expliquer par des traits généraux, par des traits contextuels, par les valeurs, par une typologie, etc.) n'a réussi à minimiser le rôle des entités qu'elle cherchait à ignorer. C'est l'une des conclusions les plus certaines d'une évaluation globale des traditions expérimentales du XX^{ème} siècle sur la personnalité. Or, souvent l'évolution rajoute des systèmes plus récents à des systèmes plus anciens dans la phylogenèse, tout en *préservant* en partie le mode de fonctionnement des derniers : le résultat est un système stratifié où plusieurs couches interagissent. La théorie de l'évolution paraît en mesure d'articuler un cadre non-réductionniste et non *ad hoc* de l'architecture cognitive complexe de la personnalité humaine. Dans le panorama théorique fragmenté de la psychologie de la personnalité, plusieurs courants sont explicitement hostiles à la psychologie évolutionniste, essentiellement à cause de préjugés et d'une interprétation simpliste des approches évolutionnistes. Par exemple, les théoriciens socio-cognitifs (nous leur consacrons un chapitre dans la deuxième partie) pensent que les hypothèses évolutionnistes ne peuvent pas rendre compte de la flexibilité de la personnalité humaine et des interactions complexes entre facteurs génétiques et influences environnementales. En réalité, la psychologie évolutionniste contemporaine est très sophistiquée et peut bien rendre compte de tout cela. En plus, elle nous offre

un cadre général pour comprendre les liens entre différents niveaux d'analyses de la personnalité, souvent étudiés par des théoriciens qui se conçoivent comme des adversaires.

En ce qui concerne le deuxième problème, nous montrerons que le domaine de la personnalité – définie comme la dimension de variation psychologique individuelle générale et reproductible – constitue une « espèce naturelle psychologique » (*psychological natural kind*), c'est-à-dire un domaine sur lequel des généralisations scientifiques intéressantes et non contingentes peuvent être formulées. Malgré les liens très forts existant entre les processus de développement de la personnalité et de fixation de l'identité subjective, l'étude de la personnalité est suffisamment autonome.

Pour résumer, le philosophe normatif devrait s'intéresser à une science de la personnalité qui étudie des dynamiques générales de variation psychologique individuelle, à plusieurs niveaux, de façon non réductionniste et dans un cadre cohérent avec la théorie de l'évolution. Une science de la personnalité n'en est qu'à ses balbutiements ; néanmoins, il n'y a pas de justification pour ignorer les généralisations intéressantes dont nous disposons déjà. La pertinence de ces connaissances pour la théorie de la vertu sera démontrée au fil des prochains chapitres.

I. Personnalité et évolution : au-delà du réductionnisme

Expliquer la variation individuelle : la sélection fluctuante

Le fait que les êtres humains ne sont pas identiques les uns aux autres en ce qui concerne plusieurs dimensions psychologiques significatives – et vraisemblablement ne diffèrent pas non plus de façon complètement chaotique – n'a pas besoin d'être rappelé. Toutefois, il faut souligner que la variation individuelle chez plusieurs espèces est un sujet très général qui fait l'objet d'une étude rigoureuse en biologie. L'existence du phénomène n'est pas anodine, et elle ne va pas sans poser de problèmes à la théorie de l'évolution. En fait, l'approche standard en psychologie évolutionniste tend à minimiser le niveau de variation intra-espèce pour des traits qui sont fonctionnellement importants. En général, s'il y a beaucoup de variation, c'est que le trait n'est pas important ; il n'avait pas de pertinence d'un point de vue adaptatif ou il ne relève pas de processus importants dans le fonctionnement de l'esprit humain. La théorie de l'évolution montre que, si un trait est associé à un avantage aussi faible qu'on veut (mais net) en termes de *fitness* – étant donné un temps suffisant (qui n'est pas forcément démesuré) – il sera amené à s'imposer inexorablement par la force des pressions sélectives. D'où l'inférence suivante : s'il y a variation, le trait concerné ne doit pas être

important. Ainsi d'après Cosmides et Tooby « (...) *la variation génétique humaine est massivement concentrée dans des différences biochimiques fonctionnellement superficielles, faisant en sorte que notre dessein fonctionnel complexe reste universel et spécifique à l'espèce*¹³⁴ ».

Cependant, plusieurs chercheurs ont récemment défendu la possibilité du maintien d'une variation individuelle significative (en termes fonctionnels et de *fitness*) par la sélection naturelle. Ce scénario est sans doute très pertinent pour le domaine de la personnalité, car il est très peu plausible que des traits tels que l'agressivité, la timidité, ou le courage n'aient pas eu de signification adaptative¹³⁵. Une solution simple serait celle d'admettre l'universalité du plan fonctionnel cognitif humain, tout en soutenant que ce plan même prévoit de complexes phénomènes d'interaction avec l'environnement. Une série de mécanismes cognitifs universels seraient déclenchés conditionnellement à la présence de certains stimuli, au déclenchement antérieur d'autres processus ou à un calibrage pendant l'enfance. Si cela devait se vérifier à une échelle suffisamment réduite (avec des processus qui répondent à un ensemble limité d'inputs), l'unicité de l'environnement et de l'histoire de vie d'un individu pourraient expliquer l'émergence d'un profil individuel unique *sans* postuler de la variation génétique¹³⁶. C'est bien là le problème. En fait, il est abondamment clair qu'il y a presque toujours une forte composante héréditaire pour la plupart des traits qui ont été jamais étudiés chez les hommes ou les animaux : les traits de personnalité, loin d'être une exception, constituent un cas typique¹³⁷. Mais si la variation individuelle *doit* être associée (au moins en partie) à la variation héréditaire, on retombe dans le problème de départ.

La solution réside dans le phénomène de la sélection fluctuante (*fluctuating selection*) : si on conçoit les traits comme étant associés en même temps à un certain nombre d'avantages et d'inconvénients (impliquant des *trade-off*) et qu'on pense qu'ils sont influencés par plusieurs gènes¹³⁸, il en résulte que le processus de sélection naturelle devra nécessairement faire des choix,

¹³⁴ J. Tooby et L. Cosmides, "The psychological foundations of culture," *The adapted mind: Evolutionary psychology and the generation of culture* (1992): 25.

¹³⁵ En ce qui concerne la personnalité, la sélection naturelle concernée est surtout (mais pas uniquement) la sélection sexuelle, plus que la sélection qui favorise des traits associés à la survie. La sélection sexuelle favorise des traits qui donnent à un individu plus de chances de succès dans la compétition sexuelle. Par définition, il doit y avoir de variation individuelle dans les traits issus de l'évolution sexuelle. Des études récentes ont confirmé que la personnalité affecte de façon décisive et prévisible le choix de partenaire. Voir G. Geher et G. Miller, *Mating Intelligence: Sex, Relationships, and the Mind's Reproductive System* (Lawrence Erlbaum, 2007).

¹³⁶ D. M. Buss, "Evolutionary Personality Psychology," *Annual Review of Psychology* 42, no. 1 (1991).

¹³⁷ Marvin Zuckerman, *Psychobiology of personality*, 2ème édition. (New York, NY: Cambridge University Press, 2005).

¹³⁸ Si plusieurs gènes sont concernés, le processus d'élimination de plusieurs mutations non optimales sera plus long.

et n'aura pas la possibilité de minimiser tous les inconvénients¹³⁹. Si l'architecture de départ (sur laquelle l'évolution travaille à un moment donné) est massivement construite sur des *trade-off* complexes, il faudra choisir un équilibre acceptable entre les avantages qu'on veut maximiser et les inconvénients qui devront être acceptés.

Même s'il en est ainsi, la théorie de l'évolution postule qu'à un moment donné et dans un environnement donné il y a un ou plusieurs équilibres optimaux dans la configuration des traits : à la fin d'une période suffisamment longue, la force inexorable des pressions sélectives devrait imposer un de ces états, et donc éliminer la variation génétique qui était à la base de chaque configuration alternative. Le problème de ce raisonnement est le suivant : dans le cas d'un ensemble complexe de traits qui influencent des aspects liés et interdépendants de la vie d'un organisme, l'équilibre optimal postulé par la théorie peut être très sensible à un changement de l'environnement (au sens large, y compris l'environnement social et la niche évolutive). Dans un contexte où l'environnement change très rapidement, l'optimum évolutif peut changer aussi, sans que la sélection puisse jamais éliminer toute la variation génétique. Les variantes qui soutiennent une configuration donnée sont favorisées lorsque cette configuration est proche de l'équilibre optimal, et défavorisées lorsqu'elle est loin de l'équilibre. Dans le territoire et dans les niches habités par l'espèce, le changement peut être suffisamment rapide pour que la configuration concernée soit favorisée un nombre de fois suffisant pour maintenir les gènes en circulation. Ce mécanisme pourrait expliquer le maintien d'une variation phénotypique héréditaire. Est-ce que le modèle s'applique à un vaste ensemble de cas connus et en particulier à celui de la personnalité humaine ?

La réponse aux deux questions semble être oui. Avant tout, comme nous l'avons dit, une variation héréditaire a été observée pour la plupart des traits, chez la majorité des animaux : « *presque tout caractère de toute espèce qui ait été étudiée intensément présente un niveau non nul d'influence héréditaire*¹⁴⁰ ». Les traits qui sont évidemment liés à la *fitness* ne font pas exception : parfois la variation prend la forme du positionnement individuel dans une dimension continue (ex : disposition plus ou moins forte à crier dans les criquets males, *Gryllus integer*)¹⁴¹, parfois on retrouve un nombre limité de « types » discrets (ex : des différents styles de cour chez le chevalier combattant¹⁴²)¹⁴³. Ensuite, d'après ce que nous savons des différentes espèces et de leurs milieux

¹³⁹ D. Nettle, "The evolution of personality variation in humans and other animals," *American Psychologist* 61, no. 6 (2006).

¹⁴⁰ M. Lynch et B. Walsh, *Genetics and Analysis of Quantitative Traits* (1998) 174.

¹⁴¹ W. H. Cade, "Alternative Male Strategies: Genetic Differences in Crickets," *Science* 212, no. 4494 (1981).

¹⁴² D. B. Lank et al., "Genetic polymorphism for alternative mating behaviour in lekking male ruff *Philomachus pugnax*," *Nature* 378, no. 6552 (1995).

naturels, il semble que les conditions requises par la sélection fluctuante devraient être réunies souvent¹⁴⁴. Un bon exemple est celui des pinsons des Galápagos¹⁴⁵ :

«...une année sèche en 1977 a réduit drastiquement la disponibilité de nourriture pour les pinsons qui se nourrissent de graines dans l'île de Daphné, favorisant les oiseaux plus grands dotés d'un bec plus profond qui pouvait ouvrir des graines plus grandes et dures. La sélection a été si forte qu'en 1978 les dimensions moyennes du corps des jeunes pinsons avaient augmenté de 0,31 écarts types par rapport à 1976, et la profondeur du bec avait augmenté de 4%. Cependant, avec le retour des pluies en 1978, les graines plus petites et douces étaient devenues à nouveau abondantes et la direction de la sélection avait changé à nouveau. La population des pinsons des Galapagos maintient un haut niveau de variation génétique grâce à cette sélection fluctuante¹⁴⁶. »

D'innombrables autres exemples existent. Les chabots vivent aussi bien dans des eaux infestées par des prédateurs que dans des zones plus sûres. Il a été démontré qu'ils varient suivant une distribution normale dans leur attitude de vigilance face aux prédateurs, et que cela est causé par les gènes (les poissons qui ne sont pas vigilants ne le deviennent pas dans des eaux infestées, les poissons vigilants ne cessent pas de l'être dans des eaux calmes). Une attitude de vigilance qui n'est pas nécessaire consomme des énergies précieuses et devient contresélectionné dans des eaux sûres¹⁴⁷. Les mésanges charbonnières varient également selon une distribution normale dans leur attitude de courage dans l'exploration. Plus de courage favorise la survie dans une période de pénurie de nourriture, mais expose à des risques inutiles dans des périodes normales¹⁴⁸.

En ce qui concerne la personnalité humaine, il est clair que la variation est largement héréditaire, due à plusieurs gènes et que l'environnement a été très variable au fil de l'évolution¹⁴⁹. Pour que le

¹⁴³ Nettle, "The evolution of personality variation in humans and other animals," 623.

¹⁴⁴ Existence de *trade-off* complexes ; influence par plusieurs gènes sur un trait ; variation rapide et continue de l'environnement (au sens large).

¹⁴⁵ B. R. Grant, *Evolutionary Dynamics of a Natural Population: the large cactus finch of the Galápagos* (University of Chicago Press, 1989) 86-89.

¹⁴⁶ Nettle, "The evolution of personality variation in humans and other animals," 624.

¹⁴⁷ S. O'Steen, A. J. Cullum et A. F. Bennett, "Rapid evolution of escape performance in Trinidadian guppies (*Poecilia reticulata*)," *Evolution* 56 (2002).

¹⁴⁸ N. J. Dingemanse et al., "Fitness consequences of avian personalities in a fluctuating environment," *Proceedings: Biological Sciences* 271, no. 1541 (2004).

¹⁴⁹ Les études des jumeaux monozygotes élevés dans des familles différentes ont montré de façon convaincante que toute variable individuelle significative (l'intelligence, les traits de personnalité, la sociabilité, la prédisposition à la satisfaction dans la vie, les styles cognitifs, etc.) est largement héritable, souvent à hauteur de 50% (voir Walter

modèle puisse être appliqué, il manque un compte-rendu des effets des différentes configurations de traits de personnalité sur la *fitness* : ce n'est pas important que l'environnement ait changé beaucoup, s'il n'a pas changé par rapport à l'effet des traits sur la *fitness*. Pour la plupart des traits, l'idée selon laquelle des niveaux différents sont plus ou moins optimaux (suivant l'environnement) paraît plausible. Si on reste dans le domaine de la spéculation, l'hypothèse contraire (il existe des niveaux *x*, *y*, *z* d'extraversion, courage et loyauté qui sont toujours optimaux) paraît difficile à accepter.

On peut considérer les exemples suivants : l'extraversion est liée à plusieurs effets positifs, dont un nombre plus élevé de partenaires sexuels (la signification adaptative est claire), une disposition plus poussée à l'infidélité et à arrêter une relation en cours pour une autre, une bonne capacité à former des relations sociales, un support social plus important et une activité physique plus soutenue¹⁵⁰. Cependant, l'extraversion semble exposer davantage au risque, et les extrovertis sont hospitalisés pour accident ou maladie plus que les autres : en général, ils ont plus de blessures traumatiques, ils sont plus disposés à des actes criminels, antisociaux et à être arrêtés. Ils sont aussi plus disposés à l'émigration¹⁵¹. Suivant que l'environnement ancestral récompense la prise de risque ou une attitude circonspecte, le niveau optimal d'extroversion peut changer. Même des traits comme le névrotisme

Mischel, *Introduction to personality : toward an integration*, 7th édition. (Hoboken, NJ: J. Wiley & Sons, 2004) 330.). Cela signifie que 50% de la variation est expliqué par des facteurs génétiques : l'environnement et les interactions gène/environnement jouent un rôle tout aussi important. En plus, le temps est révolu où les généticiens pensaient pouvoir trouver « le gène » de l'intelligence ou de la timidité : l'étude du génome humain a montré que toute variable individuelle significative est contrôlée par un réseau complexe de plusieurs gènes (voir Zuckerman, *Psychobiology of personality*). En ce qui concerne la variation de l'environnement ancestral, des données de paléoclimatologie montrent que le pléistocène a été caractérisé par une fréquence très élevée de changements climatiques (voir R. Boyd et P. J. Richerson, "Not By Genes Alone: How Culture Transformed Human Evolution," (Chicago: University of Chicago Press, 2004), 131.).

¹⁵⁰ T. Buchanan, J. A. Johnson et L. R. Goldberg, "Implementing a five-factor personality inventory for use on the internet," *European Journal of Psychological Assessment* 21, no. 2 (2005), C. Chen et al., "Population Migration and the Variation of Dopamine D4 Receptor (DRD4) Allele Frequencies Around the Globe-a family-based approach," *Evolution and Human Behavior* 20, no. 5 (1999), R. E. Franken, K. J. Gibson et P. Mohan, "Sensation seeking and disclosure to close and casual friends," *Personality and individual differences* 11, no. 8 (1990).

¹⁵¹ C. A. Field et G. O'Keefe, "Behavioral and psychological risk factors for traumatic injury," *Journal of Emergency Medicine* 26, no. 1 (2004), D. Nettle, "An evolutionary approach to the extraversion continuum," *Evolution and Human Behavior* 26, no. 4 (2005), J. Samuels et al., "Personality dimensions and criminal arrest," *Comprehensive Psychiatry* 45, no. 4 (2004).

ou l'ouverture à l'expérience ne se révèlent pas toujours (respectivement) négatifs ou positifs¹⁵². Le névrotisme est un bon indice de la plupart des troubles psychologiques, il est associé à une mauvaise condition de santé physique (probablement par le biais du stress) et généralement lié à des niveaux élevés d'émotions négatives telles que la peur, la tristesse, l'anxiété et la culpabilité. Néanmoins, chez plusieurs espèces, le névrotisme semble être corrélé à une vigilance face à des dangers extrêmes : il paraît que dans des sous-groupes de la population qui prennent des risques extrêmes (ex : alpinistes), le niveau de névrotisme est exceptionnellement bas¹⁵³. Le névrotisme semble aussi être associé au succès académique et scolaire chez les individus qui sont en mesure de tolérer ses effets négatifs, favorisant ainsi la compétitivité¹⁵⁴. Il semble que, suivant le genre et la distribution des dangers et des défis présents dans l'environnement ancestral, le niveau optimal (à un moment donné) du névrotisme a pu varier.

L'ouverture à l'expérience, caractérisée par un style cognitif alternatif et divergent – qui recherche toujours la nouveauté et établit des liens inattendus entre des domaines éloignés –, est associée à un niveau d'attractivité sexuelle plus important. Le succès en amour des créatifs (dimension essentielle de l'ouverture à l'expérience) en est la preuve¹⁵⁵. Cependant, cette non-conformité cognitive semble générer également des problèmes, car elle est associée à un risque plus élevé de psychose, schizophrénie et dépression, ainsi qu'à une tendance à la crédulité quant au paranormal¹⁵⁶. Si même des traits dont la valeur paraît si nette (positive ou négative) au premier abord imposent en réalité des *trade-off* très forts, il est clair qu'il n'y a pas une seule configuration des niveaux des traits qui

¹⁵² Le névrotisme et l'ouverture à l'expérience sont deux des cinq grandes dimensions qui définissent la personnalité humaine dans le modèle du *big five*. Ces traits correspondent à des tendances générales qui peuvent se manifester de plusieurs façons. Par exemple, le névrotisme est une tendance générale à éprouver des émotions négatives. Le but de la première partie de cette thèse est l'étude des conditions de possibilité d'un dialogue entre psychologie de la personnalité et théorie de la vertu. Toutefois, nous sommes obligés de mobiliser des notions techniques de psychologie de la personnalité: le lecteur qui n'est pas sûr de pouvoir suivre devrait lire au moins le premier chapitre de la deuxième partie (consacrée à une présentation de la psychologie de la personnalité) avant.

¹⁵³ G.M. i Freixanet, "Personality profile of subjects engaged in high physical risk sports," *Personality and Individual Differences* 12, no. 10 (1991).

¹⁵⁴ J. McKenzie, "Neuroticism and academic achievement: the furnace factor," *Personality and individual differences* 10, no. 5 (1989), J. McKenzie, M. Taghavi-Khonsary et G. Tindell, "Neuroticism and academic achievement: the Furnace Factor as a measure of academic rigour," *Personality and Individual Differences* 29, no. 1 (2000).

¹⁵⁵ D. Nettle et H. Clegg, "Schizotypy, creativity and mating success in humans," *Proceedings: Biological Sciences* 273, no. 1586 (2006).

¹⁵⁶ Nettle, "The evolution of personality variation in humans and other animals," 626.

aurait été constamment la meilleure dans l'environnement ancestral : le modèle de la sélection fluctuante s'applique parfaitement au cas de la personnalité humaine.

Personnalité animale

Peut-on dire sérieusement qu'il y a une personnalité chez les animaux ? À partir de l'expérience directe de ceux qui interagissent avec des animaux et des récits informels du comportement de différents individus dans une troupe, il est abondamment clair qu'il y a des différences individuelles stables parmi les animaux. En ce qui concerne ces données informelles, à part l'expérience directe des propriétaires de chiens, il y a eu des tentatives de description systématique et détaillée des individus (n=30) d'une troupe de chimpanzés dans leur milieu naturel et pendant une longue période¹⁵⁷. Cette description a révélé chez les chimpanzés l'existence de traits tels que la sociabilité, l'agressivité, la dominance et une disposition à la peur.

Beaucoup plus importantes sont les nombreuses études récentes qui appliquent aux animaux la même méthodologie utilisée pour élaborer les questionnaires à la base de la théorie des traits humains : l'analyse factorielle (voir ci-dessous). Pour bizarre que cela puisse paraître, il est plutôt facile de transposer la méthode de traitement des données d'un questionnaire au comportement animal : après tout, l'équivalent d'une réponse à une question peut être l'observation de la présence ou absence d'une réponse comportementale dans un contexte donné, ou l'évaluation sur une échelle de l'attitude de l'animal dans une situation. Ensuite, le traitement statistique peut être identique. Des études de ce genre ont mis en évidence l'existence de traits : une recension de 19 études chez 12 espèces a révélé que des traits comme l'agressivité, la sociabilité et la disposition à la peur sont présents dans la plupart des cas¹⁵⁸. Une deuxième recension de la littérature qui se concentrait sur 4 espèces (chat, hyène, chimpanzé et singe) a souligné l'existence de la sociabilité, le continuum dominance-soumission, l'anxiété, l'agressivité et l'attitude au jeu¹⁵⁹. Parmi les espèces qui ont été étudiées le plus, on retrouve les chiens (attitude au jeu, curiosité, sociabilité et attitude à la chasse – ce dernier étant un trait spécifique à l'espèce)¹⁶⁰, les rats¹⁶¹ et surtout les chimpanzés, à cause de

¹⁵⁷ J. Goodall, *The Chimpanzees of Gombe: Patterns of Behavior* (Belknap Press of Harvard University Press, 1986).

¹⁵⁸ S. D. Gosling et O. P. John, "Personality Dimensions in Nonhuman Animals: A Cross-Species Review," *Current Directions in Psychological Science* 8, no. 3 (1999).

¹⁵⁹ S. D. Gosling, "From mice to men: what can we learn about personality from animal research?," *Psychological Bulletin* 127, no. 1 (2001).

¹⁶⁰ K. Svartberg et B. Forkman, "Personality traits in the domestic dog (*Canis familiaris*)," *Applied animal behaviour science* 79 (2002).

leur proximité évolutive. Dans une étude systématique de 100 chimpanzés dans 12 parcs zoologiques, les chercheurs ont essayé de retrouver la structure de la personnalité humaine en cinq facteurs (le *big five*, voir deuxième partie) chez les chimpanzés. Le modèle en six facteurs (audace (*boldness*), activité, dépendance, niveau émotionnel, jovialité et ouverture) qui en est résulté n'a pas de correspondance facile avec le *big five*, mais il paraît avoir une claire signification adaptative pour la vie dans le groupe¹⁶².

Quel genre de continuité existe entre la personnalité animale et humaine ? D'après Zuckerman :

« *Il y a quatre caractéristiques fondamentales qui ont été identifiées chez des animaux non humains, en particulier ceux qui vivent dans des groupes sociaux, dans le tempérament des bébés et des enfants et dans la personnalité des adolescents et adultes* ».

Il s'agit de (respectivement : animal/enfant/adulte) la sociabilité(/sociabilité/extraversion), la prédisposition à la peur(/tempérament négatif/névrotisme), l'agressivité (colère/agressivité-jovialité) et la curiosité-exploration(/curiosité/recherche d'émotions)¹⁶³.

Il ne faut pas surestimer ces conclusions sur la personnalité animale : si la psychologie de la personnalité (relative à l'homme) n'a commencé que récemment à évoluer de façon rapide et encourageante vers la sortie d'une crise des fondements très longue, on ne peut penser que ces études sur les animaux, dont le volume est beaucoup plus faible, présentent moins de problèmes conceptuels et nous donnent une image satisfaisante. En particulier, on sait que le risque d'une compréhension de la personnalité des animaux dans leur milieu naturel est celui de l'anthropomorphisme, alors que dans le milieu contrôlé du laboratoire et de la cage on risque de produire des réponses totalement artificielles.

Cependant, si la conclusion qui paraît légitime n'est pas la célébration d'une science *mûre* de la personnalité animale et humaine, il est clair qu'il existe un domaine de phénomènes qui a son unité et importance pour des raisons évolutives, et qui attend d'être exploré par des moyens les plus rigoureux possibles. Ce serait très étonnant que – comme le soutiennent les situationnistes radicaux et d'autres sceptiques sur la notion de personnalité – l'évolution ait laissé au hasard et à l'accumulation de traits neutres un domaine aussi sensible que la variation individuelle. L'ambition

¹⁶¹ J. R. Royce, "On the construct validity of open-field measures," *Psychological Bulletin* 84, no. 6 (1977). Ils sont utilisés pour trouver des analogues à l'agressivité, l'anxiété et la recherche de sensations (*sensation seeking*) : chez les rats on a repéré des traits tels que l'activité, la réactivité émotionnelle et l'attitude au marquage du territoire.

¹⁶² J. E. King et A. J. Figueredo, "The Five-Factor Model plus Dominance in Chimpanzee Personality," *Journal of Research in Personality* 31, no. 2 (1997).

¹⁶³ Zuckerman, *Psychobiology of personality* 252.

de trouver le niveau d'analyse approprié pour des généralisations intéressantes sur la structure de la personnalité humaine paraît tout à fait justifiée¹⁶⁴.

L'évolution : un cadre méta-théorique pour la structure de la personnalité

Comme c'est le cas dans plusieurs domaines de la biologie et de la psychologie, nous pensons que la théorie de l'évolution est en mesure d'offrir un cadre conceptuel méta-théorique qui rend justice à la complexité du phénomène de la personnalité humaine. Une fois ce cadre en place, nous pourrions expliquer de façon convaincante la tendance des théories scientifiques et philosophiques de la personnalité à se concentrer sur certains niveaux d'analyse, en délaissant les autres. Cette tendance réductionniste a empêché l'émergence d'un cadre conceptuel consensuel.

L'évolution procède par bricolage. Pour modifier le système face à de nouvelles pressions sélectives, l'évolution tend à tout recycler – quitte à introduire quelques modifications – avant d'ajouter de nouveaux éléments. Ainsi, on retrouve typiquement une architecture cognitive « stratifiée ». Des systèmes cognitifs plus anciens sont maintenus dans une version modifiée, pour permettre l'interface et la coopération avec des systèmes plus récents. Les théories « duales » de la cognition constituent un bon exemple : à la cognition intuitive, automatique et fortement émotionnelle (système 1) vient s'ajouter une capacité de réflexion explicite, intentionnelle et qui demande plus d'effort (système 2). Aujourd'hui, les deux systèmes coopèrent : en ce sens, pour être rigoureux, il faudrait dire que le système 1 actuel est seulement un homologue du système intuitif et émotionnel dont nos ancêtres étaient équipés. Ce dernier n'avait pas de capacité d'interface avec un système de plus haut niveau. Malgré cela, le système 1 actuel maintient un mode de fonctionnement

¹⁶⁴ Il est vrai que les situationnistes ne soutiennent pas que l'évolution des différences individuelles a été une accumulation chaotique de traits neutres. Doris conjecture plutôt que la variation radicale du comportement humain dans chaque situation est en elle-même une adaptation, voir Doris, *Lack of character: personality and moral behavior*. L'argument touche à quelque chose de vrai, car l'évolution a besoin de réponses spécifiques à des problèmes spécifiques, ce qui va dans le sens du particularisme. Mais le situationniste radical, rejetant toute forme d'organisation de plus haut niveau dans la personnalité, est réduit à envisager l'ensemble des réponses situationnelles comme une simple liste. Il n'a pas d'explication quant à l'existence d'une certaine distribution des réactions individuelles (par rapport à une autre) dans les individus. Si l'évolution favorise des réponses spécifiques à des problèmes spécifiques, il n'est pas plausible qu'elle choisisse ou distribue au hasard la liste de réponses spécifiques que les différents individus manifestent. La charge de la preuve d'expliquer comment cela serait possible revient au situationniste. Voir le chapitre sur le situationnisme.

très proche de celui de son homologue (intuitif, automatique, émotionnel), et cela marque profondément la cognition humaine¹⁶⁵.

De la même façon, la personnalité humaine est déterminée par une série de systèmes stratifiés qui se sont ajoutés au cours de la phylogenèse, chacun introduisant une source de variation individuelle répondant à des pressions évolutives différentes. Chaque système est en interface avec les autres, mais il retient son mode de fonctionnement originaire, qui influence l'équilibre global du système. Chez l'homme, en allant du plus ancien au plus récent dans la phylogenèse, on peut envisager quatre niveaux :

(1) le tempérament : un ensemble de dispositions générales mais peu fiables, avec une base neurologique claire et des homologues chez les animaux. Par exemple, une tendance générale à l'extraversion ou au névrotisme sont générées par des systèmes neurobiologiques (respectivement) de motivation et d'inhibition, très anciens et bien étudiés chez l'homme. Des environnements sociaux et physiques différents favorisent des tempéraments différents, d'où la persistance de la variation dans le pôle génétique. L'étude du tempérament est très bien développée : la psychologie de la personnalité classique a exploré des traits de ce genre pendant près d'un siècle¹⁶⁶.

(2) les stratégies comportementales ou traits contextuels : peut-être aussi chez l'animal, mais certainement chez l'homme, il existe des traits plus spécifiques et contextuels, mais plus fiables. Par exemple, une timidité ou une anxiété sociale qui visent une liste spécifique de situations, et pas d'autres. Dans la situation concernée, l'attitude se manifeste de façon très fiable. On peut être bloqué par la peur de parler en public, mais savoir aborder une personne de l'autre sexe, ou vice versa. Ce sont autant de stratégies comportementales qui peuvent donner des avantages dans des environnements différents. L'étude de la personnalité à ce niveau d'analyse en est à ses balbutiements. Par exemple, Buss a étudié la variation individuelle au niveau de stratégies contextuelles dans le choix du partenaire, la gestion des relations et le travail dans des milieux compétitifs¹⁶⁷. Pour une présentation des théories principales qui modélisent ces traits contextuels d'un point de vue cognitif, voir le chapitre sur les approches socio-cognitives à la personnalité.

(3) les outils de gestion de la réputation : comme nous l'avons montré dans le chapitre précédent, dans toutes les cultures, il existe une série de concepts évaluatifs (trait, type) mobilisés dans le processus de définition et redéfinition de la réputation, dans un contexte social. Nous formons des

¹⁶⁵ S. Chaiken et Y. Trope, *Dual-process theories in social psychology* (Guilford Press New York, 1999).

¹⁶⁶ Zuckerman, *Psychobiology of personality*.

¹⁶⁷ D. M. Buss et D. P. Schmitt, "Sexual strategies theory: A contextual evolutionary analysis of human mating," *Psychological Review* 100 (1993), O. C. H. Lund et al., "Tactics of hierarchy negotiation," *Journal of Research in Personality* 41, no. 1 (2007).

croyances à propos de nous-mêmes et d'autrui : ainsi, certaines personnes sont considérées comme courageuses, d'autres comme loyales, ou déloyales, généreuses, timides, etc. Ces « réputations » sont « négociées », car nous luttons pour qu'une certaine image s'impose. Si, par exemple, nous réussissons à acquérir une réputation de courage (méritée ou non), cela va nous donner des avantages sociaux. De la même façon, cette réputation peut nous aider à devenir vraiment courageux, car la stabilisation et l'intériorisation d'une réputation affectent systématiquement les attitudes comportementales.

(4) le « soi » : enfin, tous ces niveaux stratifiés de variation individuelle sont mobilisés et intégrés dans un processus d'individuation qui aboutit à la définition d'une identité personnelle. La difficulté à comprendre la structure cognitive et le fonctionnement du « soi » justifie l'existence d'un nouveau domaine interdisciplinaire foisonnant : l'étude de la subjectivité (*subjectivity research*)¹⁶⁸.

Chez plusieurs animaux sociaux, la variation tempéramentale répond à des opportunités d'adaptation à un environnement changeant. Dans l'environnement social complexe de l'homme, cette logique est poussée à l'extrême. Il existe des opportunités très spécifiques et contextuelles à saisir : le fait de s'afficher ou de se tenir à l'écart de situations et de rôles sociaux précis (parler en public, prendre l'initiative dans la formation des alliances, être un leader ou un allié fiable) produit des conséquences distinctes. D'où l'intérêt de traits plus spécifiques et contextuels : par exemple, une timidité ou un courage qui visent certaines situations plus que d'autres.

Or, les traits spécifiques interagissent avec les traits généraux : avoir une propension générale à la timidité favorise sans doute l'émergence de plusieurs formes de timidité spécifique, et les manifestations de timidité « sélective » sont probablement associées à l'activation des systèmes neurobiologiques qui déterminent le tempérament. Néanmoins, comme nous le montrons dans la deuxième partie, l'existence des stratégies comportementales contextuelles n'efface pas la contribution spécifique du tempérament. D'un point de vue cognitif, il n'y a pas de simple substitution ; la timidité sélective ne prend pas la place de la timidité tempéramentale : il sera toujours possible de mettre en évidence des effets spécifiques à la deuxième. Par exemple, indépendamment du profil contextuel, une personne timide par tempérament prendra plus de temps pour se mettre dans un état d'esprit social : c'est comme si son « moteur social » avait besoin d'un échauffement plus long, en général.

¹⁶⁸ De nouveaux centres de recherche sont consacrés à ce domaine, par exemple, The Danish Center for Subjectivity Research, <http://cfs.ku.dk/>.

L'évolution travaille de façon récursive : elle recycle à nouveau un système qui avait déjà été développé grâce au recyclage. Ainsi, les mécanismes de gestion de la réputation, qui viennent s'ajouter au tempérament et aux traits contextuels, interagissent avec les deux niveaux inférieurs. Encore une fois, il n'y a pas de substitution : l'interaction va dans les deux sens. D'un côté, les tendances comportementales (générales et contextuelles) favorisent la stabilisation d'une certaine réputation chez un sujet. Quelqu'un qui n'a pas peur a plus de chances d'être perçu comme étant courageux (par soi-même et par autrui). De l'autre côté, une image de soi issue d'exigences spécifiques de gestion de la réputation peut devenir une prophétie auto-réalisatrice : je commence à me concevoir comme étant courageux et loyal et cela finit par avoir des conséquences comportementales dans le même sens.

Enfin, le « soi » peut être conceptualisé comme une dimension finale où tous ces différents niveaux interviennent et sont réorganisés. Le tempérament, les traits contextuels et l'image de soi jouent une partie importante dans la fixation de l'identité personnelle, à côté d'autres facteurs tels que la mémoire et son élaboration. Nous n'avons pas vocation à ouvrir un débat sur les bases cognitives de l'identité personnelle : il suffit de souligner que le principe de la préservation du mode de fonctionnement des systèmes plus anciens est respecté. Le « soi » ne fait pas disparaître les effets typiques du tempérament, des traits contextuels et des mécanismes de gestion de la réputation sur la cognition et le comportement. À nouveau, dans les structures cognitives stratifiées la causalité est bidirectionnelle, dans les deux sens verticaux (système ancien vers système récent, système récent vers système ancien).

Cette classification des différents niveaux de la personnalité est sans doute provisoire dans ses détails (quoique cohérente avec nos connaissances actuelles). Il faut surtout retenir l'hypothèse d'une architecture « typiquement évolutionniste » du système personnalité : chaque couche (1) a été ajoutée pour répondre à de nouvelles pressions évolutives ; (2) exploite et interagit avec les couches plus anciennes ; (3) malgré la complexification introduite, n'élimine pas le fonctionnement des couches plus anciennes et leurs effets caractéristiques sur la personnalité.

Même si, après une longue phase réductionniste, les psychologues commencent à reconnaître le caractère irréductible des différents niveaux d'analyse du système de la personnalité¹⁶⁹, à notre connaissance aucun théoricien n'a proposé de façon aussi explicite un cadre évolutionniste comme un outil de « conciliation » entre les adversaires. La raison en est l'hostilité vers la psychologie évolutionniste de certains théoriciens, malheureusement, souvent ceux qui proposent les théories les plus intéressantes, comme les psychologues socio-cognitifs (voir deuxième partie). Ces derniers

¹⁶⁹ Mischel, *Introduction to personality : toward an integration*.

pensent que les hypothèses évolutionnistes ne peuvent pas rendre compte de la flexibilité de la personnalité humaine et des interactions complexes entre facteurs génétiques et influences environnementales. Or, une approche évolutionniste raffinée admet sans problèmes la coexistence de plusieurs stratégies issues de l'évolution mais conçues pour être activées dans des conditions environnementales et sociales différentes, et aussi la possibilité de développer des traits partiellement ou complètement différents suivant les conditions rencontrées pendant le développement de l'organisme. Il s'agit du phénomène de la plasticité du phénotypique : un génotype est dit plastique s'il acquiert une morphologie différente selon l'environnement¹⁷⁰.

Malgré cette situation, il faut remarquer que le cadre évolutionniste du phénomène de la personnalité humaine « saute littéralement aux yeux » à partir d'une évaluation globale de la littérature. Il n'est pas nécessaire de forcer ou modifier les modèles théoriques existants pour les faire rentrer dans un moule. Les données sur la phylogenèse et l'ontogenèse de la personnalité montrent clairement que certains systèmes existent déjà chez les animaux (traits généraux, tempérament) et se manifestent plus tôt que les autres dans le développement de l'enfant (tempérament).

Dans le modèle présenté, toute forme de réductionnisme entre les différents niveaux est exclue. En général, il y a deux types de réductionnismes : causalité unidirectionnelle entre niveaux et substitution d'un niveau à l'autre au fil du développement du sujet. L'idée selon laquelle les stratégies comportementales dépendraient uniquement de la fixation préalable d'une image de soi (je me convaincs que je suis courageux et je le deviens), ou que, au contraire, l'image de soi correspond automatiquement aux stratégies comportementales déjà existantes (je suis courageux dans les faits, je vais finir par croire que je suis courageux) constituent des exemples de causalité

¹⁷⁰ Le génotype non plastique produit le même phénotype quel que soit l'environnement. Par exemple, l'abeille devient une travailleuse ou une reine suivant la quantité et qualité de l'apport nutritif reçu pendant la phase de larve, voir M. Mameli et P. Bateson, "Innateness and the Sciences," *Biology and Philosophy* 21, no. 2 (2006). L'épaisseur de la coquille des littorines du golfe du Maine dépend de la présence ou absence d'une espèce déterminée de crabes prédateurs : à partir du même génotype, la littorine développe une coquille plus épaisse en présence des crabes. L'évolution favorise la plasticité phénotypique lorsque l'adoption totale d'une stratégie alternative chez un individu serait trop coûteuse : le désavantage d'une littorine qui développait *toujours* une coquille très épaisse – même en absence de crabes – serait trop grande, à cause du coût métabolique très important de la coquille, voir P. David et S. Samadi, *La théorie de l'évolution* (Paris: Champus Université Flammarion, 2000). Il n'y a pas de raisons pour que l'approche évolutionniste à la cognition animale et humaine se prive de ces outils standard de la théorie néo-darwinienne.

unidirectionnelle. La substitution complète des tendances du tempérament de l'enfant par des traits contextuels chez l'adulte est un exemple de réductionnisme par substitution.

Or, la tradition philosophique et la plupart des théories psychologiques sur la personnalité ont été caractérisées par des réductionnismes de ces deux types.

Réductionnisme dans l'histoire des théories de la personnalité

Il y a deux notions de personnalité. La première se réfère à des processus qui contribuent à produire des actions ou des réactions émotionnelles. Par exemple, le courage fait que l'individu courageux fasse certaines choses (sauver un enfant en danger) et réagisse d'une façon appropriée. Il s'agit d'une notion causale. Mais la question « est-ce qu'il est courageux ? » a souvent un deuxième sens. En la posant, il se peut qu'on cherche à savoir si la personne se considère comme courageuse, ou plus probablement si la personne est perçue comme courageuse dans sa communauté. La question peut être paraphrasée de la façon suivante: « peut-on objectivement affirmer qu'il est courageux ? » ; « pourrait-on défendre et négocier ce jugement en société ? ». En tout cas, il s'agit de la stabilisation d'une certaine image, alors que la première notion se réfère à l'existence de processus psychologiques réels qui ont un pouvoir causal.

Ces deux notions sont sans doute liées, mais, dans l'histoire de la réflexion sur la personnalité, leur relation a été souvent mal comprise et mal articulée. De prime abord, leur rapport semble net et non problématique. Pourtant, il est impossible d'envisager une science de la personnalité sans les mettre en perspective correctement. Il y a deux types de liens entre la personnalité en tant que dispositions à l'action et la personnalité en tant que réputation : (1) un lien causal et (2) un lien descriptif.

(1) Les processus de personnalité peuvent être responsables de la stabilisation d'une certaine image : le fait que je sois courageux explique le fait que je sois considéré comme courageux. Un lien causal dans la même direction (entre processus et image) peut être plus indirect : je peux être plutôt lâche (ou au moins pas vraiment courageux) et croire (et faire croire) que je suis courageux pour ne pas avoir à reconnaître ma lâcheté. Ici c'est la lâcheté qui finit par favoriser une réputation de courageux. Malgré le manque de correspondance, il s'agit d'un cas dans lequel les dispositions réelles à l'action contribuent à fixer l'image d'une personne. Le lien causal peut aller aussi dans la direction inverse : il s'agit de la stabilisation de pratiques causées par l'adoption d'une certaine image, qui devient une sorte de prophétie auto-réalisatrice. Ainsi, quelqu'un pourrait finir par agir de façon plutôt courageuse pour répondre à des attentes fortes ou à un stéréotype imposé dans la communauté, si bien que le stéréotype est vraiment intériorisé et commence à décrire correctement les dynamiques psychologiques de l'individu.

Celles-ci sont les deux directions possibles du lien, mais elles peuvent coexister. Par exemple, il peut y avoir une tendance générale au courage, initialement développée sans la contribution de stéréotypes ou images imposées (elle peut être le fait d'influences génétiques ou d'expériences préalables) ; ensuite, l'attribution explicite d'une réputation de courage contribue à développer davantage l'attitude réelle face au risque. L'interaction entre les deux directions de causalité peut générer un *feedback loop* positif : une tendance naturelle au courage peut rendre l'individu plutôt réceptif à la réputation de courage, la pression sociale à son tour stabilisant la tendance de départ. *Feedback loops* et causalité bidirectionnelle entre plusieurs niveaux sont typiques dans l'architecture cognitive stratifiée issue de l'évolution.

(2) Le deuxième lien est descriptif : les attributions de traits de personnalité, même dans le contexte de la négociation d'une réputation sociale, se réfèrent à la personnalité en tant qu'ensemble de processus psychologiques. Normalement, « ce qui est affirmé » par une attribution de traits (le contenu) est l'existence des processus psychologiques correspondants. Même si la gestion d'une réputation est un phénomène qui a une signification sociale spécifique dans le fonctionnement de notre espèce – que les réputations concernées soient méritées ou non – la réputation prétend décrire des inclinations comportementales véritables.

La réflexion théorique traditionnelle s'est rendue coupable d'une simplification et d'une polarisation dans la conception de la relation entre les deux niveaux. Dans toutes les théories traditionnelles, une notion de personnalité domine l'autre, si bien qu'il y a deux interprétations de fond du caractère humain. Certains théoriciens étudient la formation de dispositions au comportement, la perception que les gens se forment de ces dispositions étant un phénomène intéressant, mais au plus complémentaire.

D'autres considèrent que le domaine d'étude de la personnalité est essentiellement la gestion de l'image que les gens finissent par se faire d'eux-mêmes, après une longue négociation collective. En ce qui concerne la scientificité de l'étude de la personnalité, cette polarisation pose un problème d'ambiguïté et de vague dans la délimitation du domaine : le théoricien de la vertu qui cherche à établir un dialogue avec la science devrait-il s'intéresser à la formation de dispositions psychologiques complexes ou à la gestion de la réputation dans les communautés humaines ?

Le choix n'est pas abstrait : il correspond à deux conceptions possibles d'une théorie normative du caractère.

Le modèle de l'excellence

La première est représentée par la tradition aristotélicienne, dans laquelle le caractère est défini par la fiabilité et la certitude des conséquences dans l'action, le modèle étant l'expertise pratique. Il

s'agit du modèle typique de la théorie de la vertu : développer une personnalité c'est comme perfectionner un art ou devenir excellent dans une activité athlétique. Tout comme un artisan ou un sportif de haut niveau travaillent pendant des années pour maximiser l'efficacité et le contrôle de leurs gestes, un soldat ou un général, après avoir reçu une formation adéquate, travaillent pendant des années sur leur caractère pour développer une détermination et un courage qui les assiste dans toute situation dangereuse, sans faille.

Cette conception de la vertu est très répandue : dans les traditions philosophiques orientales, l'accent sur la perfection pratique est encore plus fort que chez Aristote. D'après Confucius, les sages sont une véritable élite d'athlètes de la perfection morale¹⁷¹. Cette tradition a tendance à minimiser le rôle de la consolidation d'une image de soi et, en général, de l'auto-conscience de l'individu vertueux. Il ne s'agit pas forcément de nier que l'individu vertueux soit conscient de l'être¹⁷², mais cette conscience est le simple fait de la reconnaissance d'une capacité acquise. Elle ne joue pas un rôle décisif dans la définition et dans la genèse de la vertu. Or, le rôle de la fixation de l'image de soi dans la genèse de la personnalité est mis en évidence par la psychologie contemporaine¹⁷³. Il ne semble pas possible de réduire le développement de la personnalité à un simple dressage systématique qui fixerait un ensemble d'habitudes pratiques, que le sujet en soit conscient ou non. En d'autres termes, le caractère n'est pas simplement un ensemble complexe d'habitudes et de capacités pratiques.

Au fil des années, un athlète peut acquérir un ensemble extrêmement varié d'habitudes et de routines pour maintenir la forme et s'entraîner. L'hygiène alimentaire, le rythme du sommeil et la gestion du temps de repos vont être finement calibrés et vont permettre une flexibilité remarquable. Contrairement à ce qu'on pourrait croire, plus on a de bonnes habitudes plus on est flexible. Le déclenchement du comportement associé à une habitude est toujours conditionnel : même si normalement on se brosse les dents tous les matins à 8h, on ne pourra pas toujours le faire si on est en voyage ou lors d'une urgence. Les processus cognitifs qui implémentent une habitude sont toujours sensibles au contexte, au moins de façon minimale. Plus on a d'habitudes dans un domaine d'action, plus on aura besoin d'arbitrage et de coordination entre elles. Un joueur de tennis professionnel a des milliers d'automatismes profondément enracinés : dans des situations qui

¹⁷¹ Voir Hutton, "Character, Situationism, and Early Confucian Thought."

¹⁷² Parfois on arrive jusqu'à louer la naïveté et l'insouciance de l'individu vertueux, qui agirait spontanément de manière honnête, courageuse et loyale, sans se rendre compte qu'il est en train d'agir de façon exceptionnelle. Il ne saurait pas identifier ses actions spécialement honnêtes ou courageuses. Pour une discussion des approches intuitionnistes à la vertu, voir le chapitre sur la psychologie de l'expertise.

¹⁷³ Voir le chapitre sur les approches sociocognitives.

paraissent très proches, certains se déclenchent plutôt que d'autres, d'une façon extrêmement sensible au contexte.

La flexibilité et l'intelligence d'un système d'habitudes « de haut niveau » est le fond de commerce de la tradition aristotélicienne. Pourtant, cet incroyable travail de finition sur ses propres habitudes n'est pas suffisant pour définir le développement de la personnalité. Nous ne considérerions pas l'ensemble de routines et d'automatismes liés à l'activité du joueur de tennis professionnel comme étant son caractère. Le caractère est l'ensemble d'automatismes et de routines qui interviennent dans le processus d'individuation, de détermination et fixation du « soi ». On *a* des capacités pratiques alors qu'on *est* vertueux, même si la vertu, à plusieurs égards, est effectivement un ensemble d'automatismes intelligents. Mais ces automatismes influencent et sont à leur tour influencés par les processus de fixation d'une image personnelle et de stabilisation du « soi ». Ce « soi cognitif » est défini par un ensemble de processus psychologiques spécifiques : par exemple, nous nous définissons implicitement par un ensemble de traits, de représentations et de valeurs. Ce réseau de représentations a une fonction essentielle dans la stabilisation de l'identité de l'individu et influence aussi le comportement (voir le chapitre sur les approches socio-cognitives).

Dans la tradition aristotélicienne, les vertus sont un ensemble de compétences pratiques *sui generis*. Elles sont des capacités qui expriment et sont au service des valeurs fondamentales de l'individu. Cependant, chez le sage, la dimension de l'habitude prend totalement le dessus : les valeurs sont « diluées » dans un tissu d'habitudes tellement fin, que le rôle des processus psychologiques spécifiquement consacrés à la stabilisation des préférences individuelles vient à disparaître. Quoi que l'individu vertueux fasse, il aura le réflexe automatique d'agir de façon moralement bonne. C'est comme si des habitudes (on aurait envie de dire) « trop perfectionnées » prenaient en charge tout, laissant « sans travail » les autres systèmes cognitifs qui normalement déterminent la personnalité. Malgré les meilleures intentions, dans cette interprétation, il s'agit bien d'une forme de réductionnisme, par rapport à la pluralité des niveaux censés expliquer la personnalité. Normalement, le système de représentations qui définit les valeurs influence les choix de l'individu. Par contre, le sage aristotélicien n'a besoin d'aucune conscience de soi. Il est moralement bon d'une façon totalement spontanée, presque malgré lui. En ce sens, la psychologie morale de la vertu devient trop proche de la psychologie de l'expertise : d'un point de vue psychologique, l'individu vertueux est littéralement un « athlète moral ».

Le modèle des valeurs

La deuxième conception est représentée par une tradition qui identifie la personnalité à un noyau de valeurs : ici, l'enjeu du développement de la personnalité est la fixation d'un ensemble approprié de

valeurs et d'images de soi. Le passage à l'acte n'est pas optionnel : il doit suivre, mais uniquement dans une certaine mesure. L'essentiel est l'engagement, l'identification : si je ne peux renoncer à me concevoir en tant que rebelle, c'est cette attitude qui définit vraiment ma personnalité, même si je n'aurai pas toujours le courage de suivre mes instincts de révolutionnaire. Qu'on ne s'y trompe pas, même sans la rigueur presque « militaire » de l'efficacité du sage confucéen ou aristotélien, le processus de fixation des valeurs est souvent long et tourmenté : il suffit de penser au genre littéraire du roman de formation du XVIII^{ème} siècle et au parcours difficile qu'il faut suivre pour trouver sa propre vocation. Néanmoins, il est plus important de définir ses aspirations que de se donner les moyens, notamment grâce à l'acquisition des compétences pratiques de la tradition « aristotélienne ».

Le processus de recherche et de stabilisation des valeurs individuelles fait l'objet d'un intérêt qui dépasse largement les limites d'une théorie de la personnalité, au moins dans la culture occidentale. Dans le cadre plus spécifique d'une réflexion explicite sur les différences individuelles, on retrouve plusieurs traditions qui conçoivent la structure du caractère à partir des étapes du processus de fixation des valeurs. Par exemple, dans la philosophie et la littérature de la première moitié du XX^{ème} siècle, des auteurs tels que Alexander Shand, Scheler et Musil ont développé une distinction entre caractère (qui correspond aux compétences pratiques aristotélienne) et personnalité, déterminée par des sentiments stables, des préférences et une sensibilité à la valeur. Ces auteurs étaient plutôt sceptiques sur la possibilité de stabiliser le caractère et pensaient que la personnalité correspondait à une dimension plus profonde : alors que la force de la volonté et la capacité à entretenir des systèmes complexes d'habitudes efficaces sont soumises aux aléas de la vie, à la pression du contexte social, nos sentiments stables nous définissent de manière plus décisive¹⁷⁴. Dans le cadre de la psychologie de la personnalité contemporaine, plusieurs auteurs se sont concentrés sur le rôle de valeurs, de projets de vie, de buts et de défis autour desquels on construirait sa propre personnalité¹⁷⁵.

Cette conception de la personnalité semble délaisser l'efficacité pratique. Mais pourquoi renoncer à bâtir des vertus qui nous permettent de défendre nos valeurs dans le monde, une fois que ces dernières ont été fixées ? Il ne faut pas accepter l'abandon des ambitions classiques de la théorie de

¹⁷⁴ A. F. Shand, *The Foundations of Character: Being a Study of the Tendencies of the Emotions and Sentiments* (Macmillan, 1914).

¹⁷⁵ N. Cantor et J. F. Kihlstrom, *Personality and social intelligence* (Prentice-Hall), R. A. Emmons, "Motives and life goals," *Handbook of personality psychology* (1997), B. R. Little, "Personal Projects: A Rationale and Method for Investigation," *Environment and Behavior* 15, no. 3 (1983), L. A. Pervin, *Goal concepts in personality and social cognition* (Hillsdale, NJ: Erlbaum, 1989).

la vertu, à moins que ce rejet n'ait été suffisamment justifié à l'aune de nos connaissances psychologiques.

II La définition du domaine d'une science de la personnalité

L'idée d'un dialogue entre une science de la personnalité et une théorie normative du caractère vertueux rencontre, nous l'avons vu, trois types d'obstacles : (1) la relation problématique entre l'empirique et le normatif ; (2) la confusion liée à un rapport de filiation problématique avec le sens commun ; (3) la crédibilité scientifique de la psychologie de la personnalité.

D'abord, il faut que le dialogue entre la psychologie descriptive et la psychologie normative soit possible et ne soit pas fondé sur des confusions conceptuelles profondes. La pertinence des découvertes empiriques pour la psychologie morale a souvent été mise en cause par des raisons de principe : le premier chapitre était consacré à cette question.

Ensuite, une véritable science de la personnalité doit résoudre ses problèmes de filiation problématique avec le sens commun : toutes les notions qui relèvent de la personnalité sont empruntées ou au moins inspirées du sens commun. L'usage ordinaire des concepts de traits ou de type est trop polysémique, ambigu et contextuel pour que la science puisse les mobiliser sans un processus de désambiguïsation et de nettoyage conceptuel préalable. Le troisième chapitre était consacré à la première étape de ce processus : la compréhension en profondeur de l'usage des termes de traits par le sens commun. La désambiguïsation des théories psychologiques et philosophiques des traits fera l'objet de la deuxième section de cette thèse.

Finalement, après les questions préliminaires, il faut aller à la substance du problème et montrer que l'idée d'une science de la personnalité est crédible, étant donné nos connaissances psychologiques. C'est le problème que nous occuperons ici.

La variation individuelle existe et peut être sans doute étudiée. Mais en savons-nous suffisamment pour parler d'une science de la personnalité ? Nous venons de répondre par l'affirmative à partir des études sur la personnalité en neurobiologie et psychologie évolutionniste. Toutefois, il n'est pas possible de s'en tirer si facilement.

Mettre en cause la « scientificité » d'un domaine de la psychologie reconnu et institutionnalisé depuis près d'un siècle pourrait paraître arrogant et semble manifester une résistance non justifiée au dialogue avec les psychologues. Mais l'existence de données et faits avérés sur la variation psychologique individuelle est une chose (et à ce niveau, la crédibilité de la psychologie de la

personnalité n'est pas mise en doute). L'émergence de conclusions solides, profondes et partagées sur la structure de la personnalité en est une autre. Dans cette thèse, il s'agit d'inviter le philosophe normatif à s'inspirer de la psychologie scientifique de la personnalité, en un sens « ambitieux ». D'ailleurs, c'est la seule acception qui permet une confrontation intéressante. Si la science nous avait donné uniquement une longue liste de faits isolés sur la variation psychologique individuelle, sans aucune conclusion structurelle et profonde, les contraintes posées à une théorie normative du caractère seraient très limitées.

Dans le premier chapitre, nous avons déjà critiqué la rhétorique du « nier l'évidence » qu'on retrouve chez certains philosophes expérimentaux (Doris et Stich, par exemple). La psychologie philosophique normative ne peut *nier* une liste de faits : par définition, elle s'intéresse au devoir être et propose une transformation de la structure psychologique humaine actuelle. Au plus, elle peut ignorer l'enseignement profond sur la structure de l'esprit humain que ces faits recèlent. Dans le cadre d'un calcul coûts/bénéfices, cet enseignement pose des contraintes à un projet de restructuration du caractère. Sans conclusions psychologiques structurelles, il n'y a pas de menace pour l'autonomie de la philosophie normative. Pour cette raison, il faut délimiter le cadre conceptuel d'une science de la personnalité ambitieuse.

Délimitation du domaine de la psychologie de la personnalité : la différence entre la psychologie de la personnalité et l'étude de l'identité personnelle

Comment peut-on définir les contours du domaine scientifique propre à la psychologie de la personnalité? Tout problème ne vient pas du vague des notions de trait et de caractère héritées du sens commun. Il existe un véritable problème de délimitation du domaine et de ses ambitions théoriques. Nous allons défendre l'idée selon laquelle la psychologie de la personnalité, dans la phase présente de son évolution, doit accepter un partage clair des tâches entre l'étude des dynamiques générales et reproductibles de la personnalité et l'étude des dynamiques spécifiques à la fixation du soi et de l'identité personnelle. Elle doit limiter ses ambitions au premier domaine.

Chez l'homme, la variation psychologique individuelle est un phénomène tellement développé et stratifié que son étude nous pousse à faire des choix théoriques de fond. On ne peut pas procéder de façon linéaire, par une simple identification de « psychologie de la personnalité » et « psychologie de la variation individuelle ». La variation individuelle chez l'homme produit deux résultats qualitativement différents : chaque être humain n'est pas doté uniquement d'une personnalité, il est aussi une « personne » unique, il est un « soi », un sujet, il possède une identité unique.

Les notions de personnalité et de subjectivité se situent dans un continuum qui va du général au particulier. Tout ce qui relève de la personnalité garde une dimension de généralité : par définition, deux personnes peuvent avoir des personnalité semblables, ou des aspects de leur personnalité qui se ressemblent. Une dynamique de personnalité peut se reproduire chez des personnes différentes. Mais l'individuation chez l'homme se pousse beaucoup plus loin ; il n'existe pas deux personnes qui soient courageuses, humbles ou honnêtes *exactement* de la même façon (par exemple, par rapport à la liste de situations qui déclenchent une réaction courageuse ou humble). Chez chacun, ces aspects idiosyncrasiques (ou « idéographiques ») se démultiplient, jusqu'à définir le contour de l'identité individuelle : ce que chacun a d'absolument unique.

Peut-il y avoir une science qui étudie en même temps la personnalité et l'individualité ? Telle semble être l'ambition des psychologues de la personnalité. Pervin a donné la définition suivante de la personnalité :

« La personnalité est la complexe organisation de cognitions, affects, et comportements qui donne une direction et une structure (cohérence) à la vie d'une personne. Comme le corps, la personnalité est composée aussi bien de structures que de processus et reflète aussi bien la nature (gènes) que la culture (expérience). En plus, la personnalité intègre aussi bien les effets du passé, y compris la mémoire du passé, que les constructions du présent et du futur »¹⁷⁶

Dans le même esprit, Funder pense que la mission de la psychologie de la personnalité est de :

« rendre compte des structures de pensée, émotion et comportement caractéristiques d'un individu, et des mécanismes psychologiques – cachés ou non – qui sont derrière ces structures. »¹⁷⁷

De ces citations, on comprend que s'il y a une limite qui reste floue dans la définition d'une science de la personne, cela concerne sans doute ses ambitions. Le projet de découvrir que les êtres humains varient suivant des tendances définissables et mesurables (et qui ont des effets non triviaux pour la vie de l'individu et de la société) est – nous l'avons vu – conceptuellement et empiriquement solide. Mais on voit aussi que l'ambition des théoriciens ne s'arrête pas là : au moins du point de vue idéal, il y a l'ambition de découvrir tout ce qui fait en sorte qu'une personne est la personne qu'elle est. En ce sens, si on accepte la distinction entre personnalité et identité personnelle (on peut dire que deux ou plusieurs personnes ont une personnalité très semblable, mais l'identité personnelle de

¹⁷⁶ L. A. Pervin, *The science of personality* (New York: Wiley, 1996) 414.

¹⁷⁷ D. C. Funder, "Personality," *Annual Review of Psychology* 52 (2001): 198.

chaque individu est unique) il paraît que la psychologie de la personnalité n'a pas encore accepté une division des tâches claire et cherche à se rapprocher le plus possible d'une science de l'identité personnelle ou du « soi », c'est-à-dire de *tout* ce qui fait qu'une personne est exactement la personne qu'elle est.

Personnalité et individualité sont sans doute étroitement liées. En d'autres termes, il est vraisemblable que les mêmes processus psychologiques qui expliquent les dynamiques de personnalité jouent un rôle déterminant dans la formation du soi et dans la détermination de la « personne ». Nous avons vu que, dans une architecture évolutionniste stratifiée, l'individualité correspond probablement à la couche plus récente, qui interagit très fortement avec les systèmes cognitifs plus anciens dans la phylogénèse. Plusieurs théories scientifiques de la personnalité – notamment dans l'approche dite idéographique – s'éloignent d'une conception des traits de personnalité très généraliste et postulent davantage de variation individuelle dans la manifestation des traits : il y aurait, par exemple, plusieurs styles de courage plutôt différents. Ces théories amènent naturellement à concevoir un lien plus étroit entre personnalité et individualité (voir le chapitre sur les approches socio-cognitives). Dans le continuum qui va de l'universel au particulier, une personnalité formée par des traits plus spécifiques et moins répandus (la personnalité des approches idéographiques) se rapproche davantage de l'individualité. En général, il est plausible que les mêmes processus psychologiques qui font en sorte que je sois courageux (ou que j'aie un style spécifique de courage) contribuent à expliquer *certain*s aspects de mon individualité. Les études scientifiques de la personnalité et de l'individualité sont sans doute liées. Néanmoins, à l'heure actuelle, l'assimilation des deux problèmes et des deux domaines d'études ne paraît pas justifiée. La personnalité est une espèce naturelle psychologique (*natural psychological kind*), c'est-à-dire un domaine sur lequel des généralisations scientifiques intéressantes et non contingentes peuvent être formulées¹⁷⁸. À la différence des rochers qui pèsent exactement 120 kg, ou des supernovae qui ont explosé un lundi, les dynamiques de personnalité qu'on retrouve chez les individus sont liées par une histoire évolutive et des liens causaux communs. Il s'agit des différentes façons dont la variation psychologique individuelle a été sélectionnée pour répondre aux défis d'une espèce sociale comme la nôtre. Même si, dans un contexte plus large, ces processus jouent aussi d'autres fonctions dans la cognition humaine et répondent aussi à d'autres exigences, ils forment un ensemble qui a une identité cognitive et évolutive claire. La stabilisation de l'individualité vient

¹⁷⁸ Voir Edouard Machery, "Concepts Are Not a Natural Kind," *Philosophy of Science*, no. 72 (2005).

s'ajouter à cette logique, mais elle n'efface pas le mode de fonctionnement et les effets de la personnalité « générale ».

Expliquer en même temps les phénomènes de la personnalité et de l'individualité est trop ambitieux. Si on retient ce critère de succès, il faut admettre qu'il n'existe pas encore de véritable science de la personne. Refuser le statut de scientificité à tout ce que nous savons de la variation psychologique individuelle – dans sa dimension générale et reproductible – à cause d'une ambition démesurée serait un sacrifice inutile.

Il est compréhensible qu'une discipline cherche à expliquer de la façon la plus complète possible son champ de phénomènes, mais pour l'instant une science compréhensive de la personnalité et de l'identité personnelle n'est vraisemblablement qu'un idéal régulateur. À ce stade, la psychologie de la personnalité – discipline encore jeune et qui vient de traverser plusieurs crises de fondements – est désavantagée par le fait de ne pas modérer et délimiter davantage ses ambitions. Si elle ne se démarque pas d'ambitions excessives, elle compromet la reconnaissance de son importance. Tous ceux qui travaillent sur les notions de caractère et personnalité dans des traditions qui n'ont pas eu des contacts significatifs avec la psychologie empirique se sentent ainsi justifiés dans le maintien de leur isolement, car, si la psychologie de la personnalité *cherche* à étudier *tout* ce qui fait qu'une personne est ce qu'elle est, alors il est raisonnable de croire qu'une *science* de la personne n'est pas pour aujourd'hui (ni pour demain, d'ailleurs).

Les théoriciens qui travaillent dans d'autres traditions considèrent alors que la recherche conceptuellement et empiriquement rigoureuse n'est qu'une voie parmi d'autres (celle de la philosophie, du sens commun) : beaucoup de discours intéressants sur la personnalité circulent, et on n'est pas obligé d'écouter celui de la science plutôt qu'un autre. Ce genre d'attitude n'est pas justifié : si la complétude et la maturité d'une science de la personne ne sont pas pour aujourd'hui, la psychologie de la personnalité a néanmoins fait des découvertes et accumulé des connaissances sur le fonctionnement de la personnalité humaine. Ces découvertes définissent un domaine qui est une espèce psychologique naturelle (*natural psychological kind*), malgré ses liens importants avec d'autres aspects de la cognition. Il s'agit du domaine de la variation psychologique individuelle générale et reproductible. L'identité de ce domaine est définie sur des bases évolutionnistes et neurobiologiques claires. Toute approche qui s'en désintéresse ne fait qu'ignorer des enseignements décisifs sur la structure de la personnalité. Rien ne justifie cette ignorance, même si la psychologie de la personnalité n'a pas encore atteint sa maturité.

Pour décourager cette attitude, il est important de conserver un niveau d'analyse suffisamment général dans la notion de personnalité (si bien que deux personnes différentes peuvent avoir une

personnalité à peu près semblable), et de garder une distinction claire entre la personnalité d'un côté et la subjectivité, le « soi », et l'identité personnelle de l'autre.

Introduction deuxième partie

L'objectif de cette thèse est de mettre en marche une collaboration et un dialogue entre la théorie de la vertu et la psychologie de la personnalité empirique. Dans la première partie, nous avons montré qu'aucune raison méthodologique ne s'oppose à la possibilité de ce dialogue, et que, par ailleurs, du côté de la psychologie, il existe un partenaire respectable. A priori, rien ne nous assurait que la psychologie différentielle (qui étudie la variation psychologique individuelle, et non les structures universelles) ne cache pas une cacophonie de voix discordantes et des théories qui se placent à des niveaux d'analyse incomparables. Dans ce cas de figure, le potentiel méthodologique du dialogue entre philosophie et psychologie serait resté inexploité, faute de candidats adéquats et de cadres conceptuels unifiés.

En revanche, nous avons montré que, même si la psychologie de la personnalité est sans doute dans l'enfance, la personnalité est un phénomène psychologique qui a son unité et sa signification, notamment d'un point de vue évolutionnaire. Etant donné l'importance de la variation psychologique individuelle pour la fitness, il est improbable que des sources de variation hétérogènes se soient superposées de façon complètement déconnectée. Au moins dans un sens minimal, la personnalité doit former un « système ». La personnalité est une espèce psychologique naturelle.

Dans la deuxième partie, nous allons introduire les théories psychologiques de la personnalité qui vont nous permettre ensuite (dans la troisième partie) d'évaluer la psychologie de la vertu classique. Si les théoriciens de la vertu ignorent les découvertes de la psychologie, ils ne vont pas seulement perdre des informations décisives, mais aussi une source de clarification conceptuelle de leurs propos. Tout au long de la deuxième partie, nous montrerons comment les distinctions et les modèles psychologiques nous aident à clarifier – avant même de les évaluer – les tenants et les aboutissants du modèle néo-aristotélicien de la vertu.

La psychologie empirique de la personnalité du XX^{ème} siècle a été essentiellement une psychologie des traits. Dans le premier chapitre, nous allons introduire, expliquer et évaluer la pertinence de cette approche.

La réflexion philosophique sur la vertu a envisagé l'existence de traits de caractère fiables à l'échelle individuelle et qui interagissent d'une façon très riche entre eux, pour orienter conjointement le comportement et les réactions émotionnelles de l'agent. Même si, au départ, la psychologie de la personnalité partageait les mêmes ambitions (voir chapitre sur la cohérence), les

traits dont on postule l'existence aujourd'hui sont caractérisés par une cohérence inter-situationnelle faible : être extraverti revient à avoir une certaine tendance à l'extraversion, qui peut se manifester ici et là, mais pas partout et pas tout le temps. En plus, ces traits sont étudiés à l'échelle de la population et non pas de l'individu.

Il y a de bonnes raisons de craindre que, à l'échelle de la population, l'étude des corrélats et des conséquences des traits généraux cache l'existence de sous-groupes où l'effet des traits serait différent. Deux formes différentes d'extraversion n'auront peut-être pas les mêmes effets sur la santé ou la réussite au travail. En ce qui concerne l'interaction entre plusieurs aspects de la personnalité, thème central de la théorie de la vertu, la non-fiabilité de chaque trait est multipliée. L'interaction moyenne entre deux effets moyens va cacher encore plus de sous-types et de détails psychologiques que les deux phénomènes de départ. Dans des sous-groupes, il est possible qu'une certaine forme d'extraversion interagisse avec une certaine forme d'ouverture d'esprit, pour donner des résultats spécifiques : tout cela est perdu dans le modèle classique des traits.

Malgré ces critiques, nous allons résister à la tentation « éliminativiste » quant aux traits. Des théoriciens tels que Borsboom et Molenaar croient que les traits de la psychologie de la personnalité *mainstream* sont des purs artefacts statistiques et n'existent pas. Nous montrerons que ces critiques ne sont pas convaincantes et que les traits généraux trouvent leur place dans une conception stratifiée de la personnalité, cohérente d'un point de vue évolutionniste (voir chapitre sur la possibilité d'une science de la personnalité, section I). Notamment, les traits généraux sont les homologues de la « personnalité animale », alors que chez l'homme viennent s'ajouter d'autres sources de variation individuelle (stratégies contextuelles plus fiables, mécanismes de gestion de la réputation, le « soi cognitif », etc.).

Même si l'enseignement de la psychologie des traits n'est pas à rejeter, le théoricien de la vertu a intérêt à explorer les approches minoritaires. Nous commençons à le faire dans le deuxième chapitre, où nous cherchons à dresser la liste des desiderata pour une théorie de la personnalité philosophiquement pertinente. Cette théorie devrait être typologique et développementale. Elle doit être typologique car, si, pour retrouver la fiabilité, on introduit des processus psychologiques moins généraux que les traits classiques, il est vraisemblable que ces processus interagissent dans un certain nombre de configurations privilégiées, constituant ainsi des types. Ces types ne vont pas être aussi généraux que ceux qui sont mobilisés par les typologies classiques et définis par un certain nombre de traits larges. En ce sens, on parlera de typologie « molaire ».

Si on veut (1) dépasser les limites de la théorie des traits – sans tomber dans une démultiplication indéfinie et ingérables des types (typologie micro ou moléculaire) – et (2) éviter les risques d'une

typologie trop simplificatrice, il faut chercher à développer une typologie « molaire », c'est-à-dire une typologie définie par l'action de stratégies comportementales moins générales, mais plus fiables. Par exemple, l'idéal serait de découvrir et étudier expérimentalement *plusieurs* formes d'extroversion, de timidité ou d'ouverture d'esprit. L'interaction de ces « traits » plus spécifiques définirait des types humains à un niveau molaire.

Par ailleurs, contre la théorie des traits traditionnelle (qui postule la fixation de la personnalité à l'âge adulte) il a de bonnes raisons de croire que la personnalité change de façon intéressante au fil de la vie : la typologie molaire devrait être aussi développementale. Il faudrait pouvoir expliquer comment et pourquoi des types de personnalité évoluent en d'autres types. Quel potentiel d'évolution pour la timidité de quelqu'un qui a un niveau élevé de névrotisme ? Va-t-elle évoluer dans une autre forme de timidité, ou bien s'estomper ? C'est le genre de questions auxquelles une science de la personnalité devra s'attaquer.

Malheureusement, dans le panorama des approches minoritaires en psychologie de la personnalité, il n'y a aucune théorie empiriquement bien établie qui se rapproche de tous ces desiderata. Il n'y a pas de théorie qui soit en même temps typologique, développementale et molaire. En ce qui concerne le renouveau des approches typologiques, nous allons commenter les travaux de Caspi, Asendorpf et Morizot et la typologie tripartite qu'ils proposent (individus adaptés, sous-contrôlés et sur-contrôlés).

Finalement, dans le troisième chapitre, nous allons introduire l'ensemble des théories qui se rapprochent le plus des desiderata et de l'idéal de « ce que les philosophes voudraient que les psychologues étudient » quant à la personnalité humaine. En particulier, ces approches développent une typologie molaire (mais, pour l'instant, pas développementale). Il s'agit des approches sociocognitives, qui représentent l'application la plus pointue des sciences cognitives à la psychologie sociale, dans le domaine de la personnalité. L'inspiration est puisée de travaux de pionniers tels que Bandura et Mischel ; parmi les auteurs contemporains plus significatifs on retrouve Shoda, Cervone, Morf, Kuhl, Cloninger, Carver et Scheier ainsi que les travaux récents de Bandura et Mischel. Ces auteurs ont le mérite de mettre au service de l'étude de la personnalité l'ensemble des théories et des notions isolées qui se sont accumulées dans le domaine de la cognition sociale. Le problème qui se pose est celui de gérer la tour de Babel de connaissances spécifiques sur le fonctionnement cognitif de : schémas, stéréotypes, concepts de trait, standards évaluatifs, systèmes d'auto-régulation, le soi-cognitif, etc.

Le thème fondamental est la conception de la personnalité comme un phénomène *émergent*, à partir de l'interaction d'un ensemble complexe d'unités cognitives avec l'environnement. Les tendances comportementales émergent de l'activité d'un *réseau connexionniste* qui relie des « unités cognitives » entre elles et aux stimuli environnementaux. Les unités sont des structures cognitives encodées de façon stable dans le cerveau. Il s'agit de croyances, désirs, buts, standards évaluatifs, épisodes précis, représentations de soi-même ou d'autrui, etc.

Les théoriciens socio-cognitifs étudient le problème de la cohérence du comportement à partir de la notion de méta-cohérence du système de personnalité, conçu comme un réseau connexionniste. Un profil de personnalité individuel est défini par la distribution et la force des liens d'activation/inhibition entre les unités du réseau. Contrairement aux traits généraux, les profils des approches socio-cognitives (par exemple, personnalité narcissique, personnalité dépendante, sensibilité au rejet) sont associés à des conséquences comportementales de façon très fiable.

L'originalité des approches socio-cognitives tient au fait qu'elles représentent la seule tentative explicite de conceptualiser de manière rigoureuse (grâce aux moyens théoriques des sciences cognitives) de nouvelles formes de cohérence possibles et souhaitables. Il s'agit en même temps de travaux pionniers et de la seule référence existante pour une théorie de la vertu ambitieuse qui veuille avoir des bases empiriques solides.

Après avoir introduit les modèles psychologiques pertinents, dans la troisième partie de la thèse nous pourrions procéder directement à l'évaluation des ambitions de la théorie de la vertu néo-*aristotélicienne*, notamment (1) la possibilité d'atteindre l'excellence pratique dans plusieurs domaines de la vie morale, cette excellence étant caractérisée par une très grande cohérence dans la conduite ; (2) l'intégration et le support mutuel entre ces formes de cohérence. Nous montrerons que ce modèle n'est pas tenable et qu'une conception de la vertu en tant que « spécialisation morale » est préférable.

La psychologie traditionnelle des traits et son inadéquation pour évaluer la théorie de la vertu

Introduction et résumé

La notion de trait est partagée entre le sens commun et des traditions théoriques différentes. Une analyse détaillée de l'usage des termes de traits par le sens commun a fait l'objet d'une section spécifique : dans ce chapitre nous allons nous intéresser à la mobilisation de la notion de trait dans des contextes explicitement théoriques, en philosophie et en psychologie. En particulier, il s'agira surtout de présenter la structure conceptuelle de la notion de trait mobilisée dans la tradition de la théorie des traits, en psychologie de la personnalité. Cela aidera à articuler et à désambiguïser la psychologie morale de la théorie de la vertu, car, d'un point de vue technique et conceptuel, la psychologie de la personnalité est beaucoup plus précise et rigoureuse que la théorie classique de la vertu. Ce sera une phase indispensable pour procéder à un véritable dialogue avec la psychologie et à une évaluation des ambitions de la tradition philosophique.

L'approche des traits, dont le modèle le plus connu est le *big five*, est largement majoritaire en psychologie de la personnalité. 80% des efforts pour comprendre la personnalité et 100% des efforts des psychologues pour modéliser directement la vertu ont lieu dans ce cadre. Dans ce chapitre, nous montrerons que cette approche est profondément inadéquate pour reformuler et évaluer les ambitions de la théorie de la vertu. Les deux chapitres suivants seront consacrés à l'exploration du 20% des modèles minoritaires et « alternatifs », qui se révéleront beaucoup plus prometteurs.

La réflexion philosophique sur la vertu a envisagé l'existence de traits de caractère fiables à l'échelle individuelle et qui interagissent d'une façon très riche entre eux, pour orienter conjointement le comportement et les réactions émotionnelles de l'agent. Non seulement l'attitude d'un individu courageux devrait être cohérente et fiable, mais le courage interagit avec la justice ou la sagesse d'une manière précise et informative. Les traits dont la psychologie de la personnalité parle aujourd'hui sont caractérisés par une cohérence inter-situationnelle faible : être extraverti revient à avoir une certaine tendance à l'extraversion, qui peut se manifester ici et là, mais pas partout et pas tout le temps. En plus, ils sont étudiés à l'échelle de la population et non pas de l'individu. C'est comme si, pour vérifier que la taille aide les joueurs de basket à marquer plus, on étudiait les statistiques des équipes par rapport à leur taille moyenne, au lieu d'étudier directement les statistiques des joueurs individuels.

Dans cet exemple, la procédure peut marcher, car l'effet de la taille sur la performance de basket est dit « localement homogène ». Un effet est localement homogène si sa manifestation à l'échelle de chaque individu explique sa manifestation au niveau de la population : la taille favorise chaque joueur individuel et, par conséquent, l'équipe (taille moyenne) dans sa globalité. Plusieurs propriétés intéressantes ne sont pas localement homogènes : par exemple, une armée ou une association ne sont pas forcément efficaces et gagnantes parce que chaque soldat ou membre est spécialement efficace, à l'échelle individuelle. Peut-être, est-ce grâce à des institutions et à un modèle d'organisation supérieur. De la même façon, il y a de bonnes raisons de craindre que, à l'échelle de la population, l'étude des corrélats et des conséquences des traits généraux cache l'existence de sous-groupes où l'effet des traits serait différent. Deux formes différentes d'extraversion n'auront peut-être pas les mêmes effets sur la santé ou la réussite au travail. En ce qui concerne l'interaction entre plusieurs aspects de la personnalité, thème central de la théorie de la vertu, la non fiabilité de chaque trait est multipliée. L'interaction moyenne entre deux effets moyens cache encore plus de sous-types et de détails psychologiques que les phénomènes de départ. Il est possible qu'une certaine forme d'extraversion interagisse avec une certaine forme d'ouverture d'esprit, pour donner des résultats spécifiques : tout cela est perdu dans le modèle classique des traits.

Malgré ces critiques, nous résisterons à la tentation « éliminativiste » quant aux traits. Des théoriciens tels que Borsboom et Molenaar croient que les traits de la psychologie de la personnalité *mainstream* sont des purs artefacts statistiques et n'existent pas. Nous montrerons que ces critiques ne sont pas convaincantes et que les traits généraux trouvent leur place dans une conception stratifiée de la personnalité, cohérente d'un point de vue évolutionniste (voir chapitre sur la possibilité d'une science de la personnalité, section I). Notamment, les traits généraux sont les homologues de la « personnalité animale », alors que chez l'homme viennent s'ajouter d'autres sources de variation individuelle (stratégies contextuelles plus fiables, mécanismes de gestion de la réputation, le « soi »).

Si les traits généraux existent, ils ne représentent pas le niveau d'analyse le plus pertinent pour évaluer les ambitions de la théorie de la vertu : notre analyse des tentatives explicites de modélisation de la vertu par des psychologues (dans la tradition des traits) – notamment la classification VIA, qui devient de plus en plus populaire – montrera clairement ces limites.

La nécessité de désambiguïser la psychologie morale de la théorie de la vertu en tant que théorie des traits

Pour commencer, il est important de prendre au sérieux l'idée d'une *théorie* des traits. Dans la famille de concepts liés à la personnalité du sens commun (traits, types, stéréotypes, profils), on retrouve une grande variabilité contextuelle dans l'usage, d'une occasion à l'autre. Dans des contextes d'énonciation différents, le sens commun ne s'interdit une référence à des processus psychologiques non équivalents par les mêmes mots (par exemple, « courage ») : loin d'être un défaut, cette flexibilité pourrait expliquer l'efficacité des concepts ordinaires. Une théorie des traits, même si elle est inspirée par le sens commun, doit faire des affirmations informatives et non équivoques sur des processus actifs dans la psychologie réelle des gens.

Il y a eu une grande variété de conceptions de la nature et du fonctionnement du caractère vertueux. On pourrait s'attendre à ce que cette grande variété ait aidé considérablement à clarifier la structure logique d'une théorie des traits : s'il y a plusieurs propositions déterminées sur la table, on finit par comprendre les enjeux structuraux. Malheureusement, ce n'est pas le cas, la question reste enveloppée dans un brouillard conceptuel très épais.

Pour comprendre la nature du problème, il faut distinguer trois niveaux dans la formulation d'une théorie qui n'a pas été systématisée définitivement et qui est restée largement implicite dans une tradition (tel est le cas de la théorie de la vertu) : (1) une ontologie cohérente et précise des entités et processus décrits ; (2) un ensemble d'affirmations qui se répètent et reviennent souvent à propos des propriétés générales des entités et processus concernés ; (3) des affirmations de détail sur les mêmes entités et processus.

Prenons l'exemple des mythes grecs, imaginant d'y voir une théorie implicite sur les dieux grecs. D'un côté on peut retrouver l'information qu'Apollon s'est manifesté à Rhodes un matin de printemps ou qu'Athéna aurait surveillé une telle personne lors d'une occasion spécifique : ce sont des informations qui pourraient être retrouvées ou non dans d'autres versions du même mythe et qui n'ajoutent pas grand chose à une conception générale des dieux grecs (3). De l'autre côté, il y a des thèmes qui reviennent souvent, comme le fait que les dieux sont capables d'assumer l'aspect de personnes ou d'animaux pour tromper les hommes, ou qu'ils sont souvent vindicatifs. Il s'agit là de thèmes plus importants; on ne pourrait pas les supprimer sans changer de façon significative la conception générale des dieux (2). Finalement, le premier niveau (qui n'existe pas forcément pour le cas des mythes grecs) serait une énonciation explicite d'une ontologie cohérente des dieux grecs : de quel genre d'êtres s'agit-il ? Quels sont leurs pouvoirs, valeurs et objectifs généraux ? L'objectif d'une élaboration théorique serait de systématiser de façon cohérente l'information des autres

niveaux, surtout celui intermédiaire des propriétés typiques. Ainsi, une théorie explicite des dieux grecs doit rendre compte de ce qu'ils font typiquement, de façon à ce qu'il n'y ait pas d'incohérence entre leur nature et leurs actions et manifestations typiques.

Elaborer une théorie psychologique des traits est un processus parfaitement analogue à celui qu'on vient de décrire. D'un côté, il y a des questions très spécifiques sur des traits de personnalité ou des vertus (3) : est-ce que le courage était une vertu essentielle pour un soldat de la première guerre mondiale, ou valait-il mieux être patient et astucieux ? est-ce que la sincérité est une vertu d'un médecin ? quels sont les traits de personnalités décisifs pour un entrepreneur d'aujourd'hui par rapport aux années 80 ? De l'autre, il y a des propriétés plus générales : par exemple, les vertus sont censées interagir entre elles et se soutenir réciproquement. Il faut être en même temps courageux et honnête pour dire la vérité dans certains contextes ; une personne sympathique et quelqu'un de sociable mais discret ne vont pas agir de la même façon (ou faire les mêmes gaffes éventuelles) en public. Ce genre d'interaction est le pain quotidien d'une théorie des traits : en pratique, l'intérêt principal de cette théorie est de nous informer sur quel ensemble de traits en interaction est nécessaire, utile ou souhaitable pour tel résultat individuel ou collectif.

Or, on peut savoir que courage et honnêteté interagissent sans savoir exactement (avec le niveau de précision requis) quel genre d'entité, processus ou dynamique psychologique sont le courage et l'honnêteté. De la même manière, on pouvait savoir que les dieux grecs peuvent changer d'aspects sans avoir une théologie pleinement développée. Dans la troisième phase, il s'agit justement d'explicitier la nature psychologique de traits de personnalité, vertus, dynamiques dans la personnalité globale et dans le caractère vertueux, de façon à ce que le cadre global soit cohérent avec les propriétés communément admises par la tradition.

À ce point, la comparaison avec le cas des mythes grecs peut stimuler une question légitime. Pourquoi faudrait-il autant de précision ? Certes, personne ne se contenterait d'une définition trop vague des traits ou d'une simple répétition du sens commun. Mais cela n'est pas le cas de la tradition de la théorie de la vertu, riche en détails psychologiques sur le fonctionnement du caractère. Alors pourquoi vouloir plus ? Ne serait-ce aussi pédant et superflu que la tentative d'élaborer une théologie trop précise des dieux grecs ?

Il semble, dans toute explication en général, et encore plus dans le cas de la psychologie morale mobilisée par une théorie normative, qu'un excès de précision soit un vice¹⁷⁹. N'est-il pas possible que (par exemple) le courage en tant que vertu se réfère justement à un ensemble non clairement

¹⁷⁹ Aristote, *Aristote: Ethique À Nicomaque*, trans. J. Tricot (J. Vrin, 1972). 1098a

délimité de processus psychologiques ? Et si le courage – dans une définition plus stricte – ne devenait pas uniquement une notion pédante, mais aussi inefficace, son bénéfice étant lié à la pluralité de dimensions de la notion de départ ?

Il est important de répondre à cette préoccupation, et la réponse sera nette. Le niveau de précision psychologique qu'on cherchera à atteindre n'est pas dicté de façon arbitraire, ou par des choix personnels. Ce sera le niveau nécessaire pour vérifier que la vertu est bien associée aux bénéfices qui ont été promis par la tradition. La notion de vertu est constamment associée à la poursuite du bonheur individuel et à la possibilité d'une vie communautaire saine et épanouie. Un théoricien de la vertu devrait être en mesure de montrer que ce lien est défendable.

Même pour les versions de la théorie où la vertu n'est pas simplement associée mais *identifiée* au bonheur individuel ou collectif (dans ce cas, on n'a pas à démontrer sa capacité à *produire* quelque chose de différent d'elle-même), la nature et les implications psychologiques de la vertu doivent résulter suffisamment claires pour que l'identification soit évaluée (pour pouvoir évaluer l'hypothèse selon laquelle un ensemble complexe de phénomènes psychologiques *constitue* le bonheur). Le niveau de clarté conceptuelle et précision psychologique nécessaires pour que la théorie de la vertu devienne évaluable est *beaucoup plus élevé qu'on ne le croit* (la charge de la preuve nous revient, bien sûr) : il ne sera jamais question d'exiger une précision supérieure à celle requise si on prend au sérieux les ambitions perfectionnistes de la théorie de la vertu.

Dans ce processus, la source de clarification sera la psychologie de la personnalité. L'évolution de ce domaine est très instructive pour la théorie de la vertu. Comme nous le montrerons, les théories des traits en psychologie de la personnalité ont cédé aux mêmes tentations et connu les mêmes problèmes que la théorie de la vertu, mais elles ont su se rattraper plus vite, grâce à un processus d'auto-critique progressive de plus en plus marquée. On assiste aujourd'hui à une re-conceptualisation très profonde de la structure de la personnalité, dont la théorie de la vertu ferait bien de profiter. Pour l'instant, cette dernière s'est limitée à une adoption sélective et à une critique locale des legs de sa psychologie traditionnelle du caractère vertueux, sans engager une critique radicale de ses fondements.

Etant donné que la liste des exploits dont la vertu est censée être capable (le niveau intermédiaire d'élaboration théorique) est plutôt claire et fait l'objet d'un certain consensus, les problèmes empirico-conceptuels les plus graves des théories traditionnelles des traits se situent au niveau de la systématisation explicite. Il n'y a pas d'analyse suffisamment cohérente de la nature des traits, de la façon dont ils sont censés interagir avec d'autres traits pour manifester ses bénéfices.

La notion de trait en psychologie

Comme nous l'avons montré, si par personnalité, au sens le plus large possible, on entend la dimension globale qui caractérise la spécificité d'un individu d'un point de vue psychologique, alors une théorie de la personnalité ne doit pas être forcément une théorie des traits. En ce sens, la définition de « personnalité » formulée par Allport paraît encore adéquate : « *la personnalité est l'organisation dynamique, interne à l'individu, des systèmes psychophysiques qui déterminent son adaptation unique à l'environnement*¹⁸⁰ ». Nous avons montré qu'il y a toute sorte de dynamiques qui ne correspondent pas à l'action d'un trait mais dont on peut envisager le rôle et l'importance dans l'économie d'une personnalité.

Cependant, l'élaboration théorique sur la structure de la personnalité a été à plusieurs reprises, à des moments différents de l'histoire, influencée par la structure superficielle des expressions du sens commun, qui semble mentionner des phénomènes psychologiques discrets (le courage, la sensibilité, l'arrogance etc.) : les traits. En ce sens, les théories de la personnalité ont été de façon très prépondérante des théories des traits, beaucoup plus que cela n'aurait été nécessaire d'un point de vue strictement théorique. C'est bien le cas de la psychologie de la personnalité, née au début du vingtième siècle comme une forme de psychologie des traits. Alors que le travail empirique sérieux et la récolte de données ont commencé bien avant – lors des premiers efforts de classement et sélection des soldats et du personnel par le biais de questionnaires aux Etats-Unis, dans le cadre du premier conflit mondial – la date de naissance de la psychologie de la personnalité en tant que domaine unitaire et reconnu coïncide avec le travail de Gordon Allport, en particulier son oeuvre majeur de 1937 *Personality ; a Psychological Interpretation*. Comme nous le verrons à plusieurs reprises, les choix théoriques du fondateur du domaine vont conditionner de façon très profonde le développement de la psychologie de la personnalité, jusqu'à nos jours. Parmi ces choix, le plus important est sans doute le rôle central accordé à la notion de trait.

Dans les décennies précédant le travail de Allport, les notions relevant du thème de la personnalité n'étaient pas absentes de la scène publique ou du discours théorique, loin s'en faut. Le problème de l'époque était au contraire de mettre de l'ordre dans une prolifération de discours et notions vagues sur les fondements psychologiques de la variation individuelle. Un ensemble foisonnant de significations différentes étaient attribuées suivant le contexte à des notions telles que tempérament, caractère, trempe, etc. D'ailleurs, du point de vue des critères exigeants de la psychologie expérimentale, le sens commun n'était pas le seul incriminé. Si dans les arts en général, et dans la

¹⁸⁰ Gordon W. Allport, *Personality; a Psychological Interpretation* (New York,: H. Holt and Company, 1937).

littérature en particulier, la personnalité est analysée avec beaucoup plus de sophistication et richesse – si bien qu'on peut attribuer à un grand écrivain la maîtrise de l'art de la compréhension d'autrui – cela ne signifie pas que ces moyens plus sophistiqués soient susceptibles d'une systématisation théorique plus facile que dans le cas du sens commun. Dans une certaine mesure, il s'agit du même problème qu'on retrouve encore aujourd'hui : d'un côté, une expertise enviable dans la compréhension d'autrui et une masse grouillante d'expressions, concepts, stratégies cognitives et inférencielles consacrés à la tâche ; de l'autre, une confusion et un manque de précision remarquables dans la tentative de systématisation psychologique sérieuse de tout cela.

Deux facteurs rendaient le problème plus urgent au début du siècle qu'il ne l'est aujourd'hui : (1) au bout d'un siècle de recherche et d'efforts, quelques étapes ont été franchies dans la direction d'une conceptualisation rigoureuse de la structure de la personnalité ; (2) nous vivons dans une époque où l'assignation des individus à des types ou à des profils de personnalité est plutôt controversée : plusieurs voix se lèvent en défense de principes « situationnistes » d'influence de l'environnement et de conditionnement social, pour justifier et excuser les choix des individus. Autrefois, il était beaucoup plus courant d'expliquer le comportement et d'attribuer la faute à partir de la trempe, du genre de personne qu'on est supposé être. Au 19^{ème} siècle il était très peu commun de justifier des actes criminels, ou même des grèves, à partir des conditions de vie difficiles des responsables : l'inférence à l'existence de personnes de « mauvais caractère » était automatique¹⁸¹.

Dans cette conjoncture historique, où le besoin de clarté conceptuelle était plus pressant que jamais, la notion de trait a paru fournir une façon élégante d'encadrer et systématiser la famille de phénomènes relevant de la personnalité. Ainsi, d'après Allport, les traits constituent l'unité scientifique fondamentale de mesure de la personnalité. Ce choix théorique a été pris résolument, par toute la discipline : parmi d'autres figures clés de l'essor de la psychologie de la personnalité, Eysenk pense que la personnalité est « *une combinaison de traits que la personne manifeste dans différentes situations et qui demeure stable dans le temps* » ; d'après Guilford la personnalité est « *la configuration unique des traits de la personne*¹⁸² ». Mais quelle est la conception des traits impliquée dans cette tentative de systématisation ? À un niveau de conception générale, avant d'entrer dans les aspects opérationnels, un trait est d'après Allport un « *système neuropsychique généralisé et focalisé ayant la capacité de rendre plusieurs stimuli fonctionnellement équivalents et d'activer et guider de façon cohérente (ou équivalente) plusieurs formes de comportement adaptatif*

¹⁸¹ Craig A. Cunningham, "A Certain and Reasoned Art," in *Character Psychology and Character Education*, ed. Daniel K. Lapsley and F. Clark Power (Notre Dame, Indiana: University of Notre Dame Press, 2005).

¹⁸² H. J. Eysenck, *The Structure of Human Personality* (Methuen, 1970), J. P. Guilford, *Personality* (McGraw-Hill New York, 1959).

et expressif¹⁸³ ». Selon Cattell, les traits sont « *des structures mentales héréditaires qui expliquent le comportement et la cohérence du comportement dans différentes situations*¹⁸⁴ ». Plus récemment, Tellegen (1991) a proposé qu'un trait est une structure « organismique » (i.e., psychologique et biologique) inférée, relativement durable, qui sous-tend une famille de prédispositions cognitives, émotives et comportementales¹⁸⁵.

Au sens minimal, « les traits représentent la covariation systématique de plusieurs réponses cognitives, émotives et comportementales habituelles dans diverses situations. C'est cette covariation qui constitue l'objet de la mesure et elle est généralement qualifiée d'expression phénotypique des traits¹⁸⁶. » La notion de trait désigne le fait qu'un ensemble de comportements, réactions et activations de structures psychologiques covarient systématiquement, d'une certaine façon spécifiée. Au sens rigoureux, l'expression phénotypique d'un trait ne doit pas être associée uniquement à la manifestation comportementale : les processus neurologiques, physiologiques et hormonaux font pleinement partie du trait. Ils ne constituent pas une espèce d'antichambre causale inintéressante, mais qui précéderait l'activation véritable du trait. En ce sens, une expression phénotypique d'un trait est l'activation locale d'une partie quelconque du réseau de covariations qui, dans sa globalité, constitue le trait. À ce niveau de généralité, le statut ontologique des traits n'a rien d'original et ne devrait pas susciter des suspicions particulières¹⁸⁷.

¹⁸³ Allport, *Personality; a Psychological Interpretation*.

¹⁸⁴ R. B. Cattell, *The Scientific Analysis of Personality* (Transaction Publishers, 2008).

¹⁸⁵ A. Tellegen, "Personality Traits: Issues of Definition, Evidence, and Assessment," *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* 2 (1991).

¹⁸⁶ Julien Morizot, "Le Développement De La Personnalité De L'homme De L'adolescence Au Milieu De La Vie : Approches Centrées Sur Les Variables Et Sur Les Personnes" (2003) 7.

¹⁸⁷ Une critique très fréquente de l'existence des traits de caractère propose d'éliminer la notion de trait en faveur d'une mobilisation directe de l'expression phénotypique : quel travail explicatif sérieux accomplit la notion de timidité, si à chaque fois on pourrait éliminer toute référence à cette notion générale, en explicitant directement qu'une personne « se sent en embarras dans la situation *x*, face à un stimulus *y* et à cause de sa croyance que *z* ? » Le fait qu'un trait ne soit ontologiquement *rien d'autre* que l'ensemble de ses manifestations possibles justifie l'élimination de la notion même de trait, jugée inutile et épiphénoménale. Néanmoins, faute de raisons supplémentaires, cette critique n'est pas convaincante. On va analyser ci-dessous la question du réalisme ou anti-réalisme quant aux traits, mais on peut déjà remarquer qu'il y a une série de phénomènes qui se réduisent à la covariation de processus plus basiques, et auxquels on ne songe pas de retirer le « statut de réalité » et de pouvoir causal. Il suffit de penser à des syndromes comme la dépression ou à des processus comme le vieillissement. Le vieillissement est l'effet émergent d'une série de processus de décadence qui pourraient être cités directement, un par un. Ce n'est pas pour autant qu'il est considéré comme un processus non réel et non causal.

Une question fondamentale dans le traitement de la notion de trait chez les psychologues est la structure hiérarchique des traits et la façon dont cette structure est dérivée et étudiée. L'intuition de base est simple : il s'agit d'organiser sur une série de niveaux (et de façon récursive) les différents éléments qui covarient dans le fonctionnement psychologique de l'individu, depuis les attitudes et les comportements les plus spécifiques jusqu'aux tendances les plus générales. Un trait *x*, à un certain niveau d'abstraction, représente la covariation d'un ensemble de traits de niveau inférieur, alors qu'un trait *y* de niveau supérieur exprime la covariation de *x* et d'autres traits du même niveau¹⁸⁸. Ainsi, partant du niveau plus bas et concret, on retrouve les *réponses spécifiques* : « *Ces réponses correspondent aux pensées, aux émotions ou aux comportements spécifiques émis dans un contexte physique ou social particulier*¹⁸⁹ ». L'observation directe et systématique dans des contextes naturels ou expérimentaux est le moyen nécessaire pour mesurer ces attitudes spécifiques. À un niveau supérieur, on trouve les *réponses habituelles*, constituées par la covariation et co-occurrence de plusieurs réponses spécifiques dans des contextes différents et chez le même individu. Ce niveau d'organisation correspond aux questions qui composent les tests de personnalité : une question est une tentative de vérification directe de la présence ou absence d'une réponse habituelle, qu'elle soit cognitive, émotionnelle ou comportementale. On ne commence à parler véritablement de traits que lorsqu'on considère la covariation des réponses habituelles : en d'autres termes, la covariation des différentes réponses aux items d'un questionnaire. Ici on parle de *trait de premier niveau* ou *facette*. À partir des facettes, suivant l'approche du théoricien, on peut envisager des traits d'un ou plusieurs niveaux supérieurs, toujours à conceptualiser comme la covariation systématique des traits du niveau plus bas. Au stade où le processus s'arrête, on parle de *dimensions*, *super-traits*, *traits d'ordre supérieur* ou *super-facteurs*.

À chaque niveau, la procédure classique pour déterminer ce qui co-varie avec quoi est l'analyse factorielle, dont le but est d'isoler, dans un ensemble d'éléments non classés, des sous-ensembles dont les éléments « co-occurrent » plus qu'ils ne le font avec les éléments des autres sous-

¹⁸⁸ P. T. Costa and R. R. McCrae, "Normal Personality Assessment in Clinical Practice: The Neo Personality Inventory," *Psychological Assessment* 4, no. 1 (1992), L. R. Goldberg, "The Structure of Phenotypic Personality Traits," *Am Psychol* 48, no. 1 (1993), O. P. John and S. Srivastava, "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives," *Handbook of personality: Theory and research* 2 (1999), S. V. Paunonen, "Hierarchical Organization of Personality and Prediction of Behavior," *Journal of Personality and Social Psychology* 74, no. 2 (1998).

¹⁸⁹ Morizot, "Le Développement De La Personnalité De L'homme De L'adolescence Au Milieu De La Vie : Approches Centrées Sur Les Variables Et Sur Les Personnes" 8.

ensembles¹⁹⁰. Dans la grande majorité du travail qui a été fait en psychologie de la personnalité, la dépendance exclusive de l'analyse factorielle est si forte qu'elle constitue une faiblesse considérable, étant donné que la méthode présente ses inconvénients. Au-delà de certains choix techniques qu'il faut faire et qu'il est possible à chaque fois de contester lorsqu'on tente une analyse factorielle (choisir le type de matrice de corrélations, la méthode d'extraction et la méthode de rotation), le problème de la procédure est qu'elle est très sensible au choix des items dans les questionnaires. Si l'objectif est de dériver la structure objective de la personnalité en termes de traits, cette méthode de classement hiérarchique n'a pas les moyens de produire des résultats meilleurs que la liste de réponses habituelles sur laquelle elle est appliquée. Elle ne peut pas remédier à une sélection peu représentative de la globalité et complexité de la personnalité humaine, au niveau du questionnaire. L'argument central en faveur de la validité de l'analyse factorielle se base sur la possibilité de retrouver à peu près la même structure hiérarchique des traits dans plusieurs contextes, grâce à des questionnaires en plusieurs langues et adaptés à de différentes cultures. En particulier, les études inter-culturelles employant l'analyse factorielle semblent retrouver systématiquement les mêmes traits d'ordre supérieur. Il y a deux modèles qu'il paraît facile de retrouver partout : le *big five* (modèle à cinq super-facteurs) et le *big three* (modèle à trois super-facteurs).

Le *big five* est le dernier résultat d'une longue tradition de recherche, qui associe à l'adoption de l'analyse factorielle une stratégie lexicale pour la sélection des items. L'intuition est la suivante : les habitudes, les réponses cognitives et émotionnelles qui caractérisent la vie de l'homme et son développement individuel doivent forcément trouver leur place dans le langage. En particulier, on peut en retrouver les marques dans les termes de traits, c'est-à-dire les adjectifs qui qualifient une personne, tels que « loyal », « honnête », « timide », « courageux », « mélancolique », etc. , le présumé étant que tous les traits ont laissé des traces dans le lexique du langage. Cette stratégie, aussi ancienne que la psychologie de la personnalité, au début prévoyait littéralement des questionnaires qui évaluaient une personne par rapport à une longue liste d'adjectifs (loyal de 0 à 7, timide de 0 à 7, etc.). Cette liste était une sélection très réduite de l'ensemble des adjectifs de trait présents dans le dictionnaire. Étant donné qu'au bout de plusieurs décennies de recherche avec cette

¹⁹⁰ Pour une présentation de l'analyse factorielle, voir J. F. Finch and S. G. West, "The Investigation of Personality Structure: Statistical Models," *Journal of Research in Personality* 31, no. 4 (1997), R. P. Macdonald, *Test Theory: A Unified Treatment* (Lawrence Erlbaum Associates, 1999), J. C. Nunnally, "Psychometric Theory," *New York: McGraw-Hill* (1994).

procédure, cinq super-facteurs émergeaient constamment, il a été possible d'élaborer des questionnaires « équivalents » aux tests lexicaux et censés mesurer les mêmes traits¹⁹¹.

C'est dans cette forme de questionnaire non explicitement lexical que se déroule aujourd'hui la recherche sur le *big five*. Dans le modèle à cinq facteurs les traits de niveau supérieur (et les traits de premier ordre qui leur sont associés) sont le Névrotisme (anxiété, dépression, vulnérabilité affective, hostilité, impulsivité et faible estime de soi), la Complaisance (honnêteté, altruisme, empathie, conciliation, modestie et crédulité), la Contrainte (attention, organisation, fiabilité, dévouement, compétence, et ambition), l'Extraversion (sociabilité, confiance en soi, expressivité, énergie, activité et recherche d'excitation), et enfin l'Ouverture à l'expérience (imagination, sensibilité, esthétisme, curiosité, recherche de la nouveauté et tolérance aux différences).

Le modèle à trois facteurs a été développé avec davantage d'attention à des problèmes d'adaptation psychologique et au substrat neurophysiologique des traits, sans qu'un privilège essentiel soit donné au langage¹⁹². Dans ce modèle, on retrouve l'Extraversion (ou l'Émotivité Positive) et le Névrotisme (ou l'Émotivité Négative). Le troisième trait, la Désinhibition (ou le Psychotisme, la Contrainte), semble regrouper des aspects de la Complaisance et de la Contrainte du modèle en cinq facteurs. On considère souvent que le modèle à cinq facteurs est à peu près une extension du modèle à trois, ce qui (étant donnée la différence de méthodologie) semble montrer l'universalité de la structure mise en évidence (au moins en ce qui concerne les trois traits partagés, dont deux, névrotisme et extraversion, sont presque identiques dans les deux modèles). Or, le fait que les

¹⁹¹ Comme on peut l'imaginer facilement, la réduction à plusieurs dizaines d'une liste des termes de traits de l'ordre des dizaines de milliers ne va pas sans choix arbitraires, qui ont été souvent contestés et se sont accumulés au fil du temps, étant donné que de nouveaux théoriciens travaillaient souvent avec des vieilles « sélections », remontant très rarement au dictionnaire, voir J. Block, "A Contrarian View of the Five-Factor Approach to Personality Description," *Psychol Bull* 117, no. 2 (1995). Naturellement, ce problème se transmet au big five dans sa formulation contemporaine, car les tests ont été formulés pour décrire les mêmes traits qui avaient été isolés auparavant.

¹⁹² Ce modèle en trois facteurs a été popularisé plus particulièrement par L. A. Clark and D. Watson, "The General Temperament Survey," *Unpublished manuscript, University of Iowa, Iowa City, IA* (1990), L. A. Clark and D. Watson, "Temperament: A New Paradigm for Trait Psychology," *Handbook of personality: Theory and research* (1999), C. R. Cloninger, "A Systematic Method for Clinical Description and Classification of Personality Variants. A Proposal," *Archives of General Psychiatry* 44, no. 6 (1987), H. J. Eysenck, *A Model for Personality* (Springer, 1981), Eysenck, *The Structure of Human Personality*, H. J. Eysenck and Michael W. Eysenck, *Personality and Individual Differences : A Natural Science Approach, Perspectives on Individual Differences* (New York: Plenum Press, 1985), A. Tellegen and N. G. Waller, "Exploring Personality through Test Construction: Development of the Multidimensional Personality Questionnaire," *Personality measures: Development and evaluation* 1 (1994).

structures de personnalité du *big five* et *big three* soient retrouvées systématiquement dans d'autres cultures et d'autres langues représente certainement un résultat remarquable, qui ne peut être expliqué uniquement par les biais de l'analyse factorielle. Les problèmes des choix techniques contestables dans l'application de l'analyse factorielle et dans la sélection des items de départ ne nous donnent pas des raisons suffisantes pour considérer ce type de résultat comme un artefact. Mais l'ambition d'avoir découvert la structure universelle de la personnalité humaine n'est pas pour autant justifiée. On peut citer au moins les problèmes suivants :

Manque de complétude : il semble que, au plus, on a découvert uniquement le noyau de la structure compréhensive de la personnalité. Le résultat impressionnant est celui de retrouver la même structure dans des contextes culturels et linguistiques différents. Mais si la coïncidence est flagrante pour les deux traits de l'extraversion et du névrotisme, déjà pour la contrainte et la complaisance, la correspondance est moins nette, jusqu'au cas de l'ouverture à l'expérience, qui n'a pas été retrouvée dans plusieurs études. Il semble également qu'il y a peut-être d'autres traits à découvrir¹⁹³.

Identification douteuse entre les facteurs de tests différents : on ne retrouve qu'à *peu près* les mêmes facteurs. Cela signifie que, dans certains cas, quelques items et même quelques facettes (traits de premier niveau) ne sont pas les mêmes, si bien que le super-facteur ainsi défini est seulement *en gros* le même que celui d'origine. Si cela d'un côté paraît compréhensible (comment espérer trouver exactement les mêmes traits dans des cultures différentes ?), de l'autre nous pousse à douter de l'intérêt du résultat. Quel est le vrai intérêt de soutenir que l'extraversion est une dimension universelle du caractère humain, si de différentes versions du super-trait ne partagent pas une ou plusieurs facettes (sur cinq ou six) ?

Traits spécifiques à des cultures : la spécificité culturelle ne se manifeste pas uniquement dans la coïncidence imparfaite entre les super-traits, mais aussi dans l'existence de traits spécifiques à des cultures particulières, qu'on ne retrouve pas ailleurs. Par exemple, Bond a isolé un trait spécifique à la culture chinoise : l'attitude envers la tradition¹⁹⁴.

¹⁹³ M. C. Ashton and K. Lee, "A Theoretical Basis for the Major Dimensions of Personality," *European Journal of Personality* 15, no. 5 (2001), S. V. Paunonen and D. N. Jackson, "What Is Beyond the Big Five? Plenty!," *Journal of Personality* 68, no. 5 (2000), Tellegen and Waller, "Exploring Personality through Test Construction: Development of the Multidimensional Personality Questionnaire.," N. G. Waller, "Evaluating the Structure of Personality," *Personality and psychopathology* (1999).

¹⁹⁴ M. H. Bond, "Localizing the Imperial Outreach: The Big Five and More in Chinese Culture," *American Behavioral Scientist* 44, no. 1 (2000), A. T. Church, "Culture and Personality: Toward an Integrated Cultural Trait Psychology," *Journal of Personality* 68, no. 4 (2000), A. T. Church, "Personality Measurement in Cross-Cultural Perspective," *Journal of Personality* 69, no. 6 (2001).

Pour résumer, la situation actuelle de la théorie des traits en psychologie et celle d'une tradition de recherche qui a développé des résultats fort intéressants, difficilement congédiables comme des artefacts (comme certains critiques radicaux le voudraient, voir ci-dessous), mais aussi difficiles à interpréter. Il paraît en effet plutôt douteux qu'ils puissent justifier l'ambition d'avoir découvert les structures fondamentales et universelles de l'organisation de la personnalité humaine.

Relation entre psychologie et théorie de la vertu

En ce qui concerne le lien entre la psychologie de la personnalité et l'éthique de la vertu, le problème de l'universalité – largement débattu en psychologie – n'est pas le plus pressant. Même si la structure du *big five* était parfaitement universelle, encore faudrait-il s'interroger sérieusement sur le type de traits postulés par les psychologues et ses liens avec la notion traditionnelle de vertu. Car le succès (relatif) de la psychologie de la personnalité vient s'ajouter au climat de respect et d'acceptation générale de la notion de trait dans le sens commun et d'autres contextes non-théoriques, résultant en une consécration ultérieure (s'il en fallait) qui ne peut que rassurer tout théoricien qui, comme le philosophe de la vertu, présuppose cette notion dans sa production théorique. En d'autres termes, si la notion de trait n'a jamais fait l'objet d'une méfiance répandue, *a fortiori* on a encore moins à craindre aujourd'hui : un consensus règne parmi les psychologues qui consacrent des carrières entières à étudier l'influence de tel ou tel autre trait sur un aspect spécifique de la vie.

Cette image est très éloignée de la réalité des choses. La notion de trait qui a trouvé sa consécration en psychologie n'a rien à voir avec la notion de vertu postulée par les philosophes, et son succès ne recèle que de mauvaises nouvelles pour la théorie de la vertu traditionnelle. Deux problèmes structuraux très profonds empêchent le philosophe de profiter du succès de l'approche des traits en psychologie : (1) le manque d'« homogénéité locale » (voir ci-dessous) dans les phénomènes étudiés par la psychologie de la personnalité, ce qui fait qu'on décrit le fonctionnement des traits uniquement au niveau de la population, plus exactement par rapport à l'individu moyen ; (2) le fait que les psychologues se réfèrent à des traits qui ont une cohérence inter-situationnelle beaucoup plus faible que celle envisagée par les philosophes (et c'est bien ce niveau très élevé de cohérence qui est responsable des bénéfices attribués à la vertu). Cela signifie que la psychologie des traits est très mal placée pour étudier l'interaction des traits à niveau individuel, et même si des développements méthodologiques permettaient de résoudre ce problème (ce qui est possible, envisageable et en train d'être fait) la notion de trait concernée resterait sans intérêt pour le philosophe de la vertu traditionnel.

Comme nous l'expliquerons dans le chapitre consacré à la recherche de la cohérence, il est vrai qu'au début, les psychologues avaient des ambitions très poussées sur la cohérence inter-situationnelle des traits, de la même teneur que les philosophes. Mais, au bout d'un parcours long et tourmenté, lors du nouveau consensus atteint à la fin des années 80 après la crise situationniste longue de deux décennies, l'exigence d'une cohérence très forte a été explicitement sacrifiée. C'est après ce tournant que se placent le succès et la consécration actuels du modèle des traits. Il est donc hors de question d'attribuer *aujourd'hui* à la psychologie une prétention à la cohérence inter-situationnelle comparable à celle des philosophes.

Le problème de l'homogénéité locale et l'interprétation réaliste des traits de caractère dans la tradition philosophique et psychologique

L'un des problèmes les plus sérieux des théories traditionnelles des traits et du caractère vertueux est l'idée selon laquelle il existe une notion unitaire de trait (ou de vertu) responsable des bénéfices et de l'épanouissement aussi bien au niveau individuel que social. Grâce aux travaux d'éclaircissement conceptuel et empirique de la psychologie de la personnalité, on peut montrer qu'on a affaire au contraire à deux types de traits très différents : voyons pourquoi à partir du cas de la théorie psychologique des traits.

La plupart des travaux qui constituent le corpus des connaissances empiriques de la psychologie de la personnalité consistent à essayer de postuler et valider l'existence d'un certain nombre de traits, et d'en étudier ensuite les relations et l'interaction. Cette étude se développe au niveau de populations, de groupes. Mais les conclusions sont censées s'appliquer au niveau de processus causaux actifs dans la psychologie de l'individu. Cela pose le problème qui est appelé en jargon technique d'« homogénéité locale »¹⁹⁵. Un processus est localement homogène lorsque les propriétés causales actives au niveau de la population et celles actives au niveau de chacun des individus dans cette population sont les mêmes. Un exemple : imaginons que l'on fasse une étude sur la corrélation entre la taille et les paniers marqués au basket, et imaginons que le résultat soit une forte corrélation, statistiquement significative. Qu'a-t-on appris par là ? Formellement, les moyens statistiques employés nous autorisent à tirer uniquement la conclusion suivante : si on prend deux *populations*, celle des joueurs très grands et celle des joueurs très petits, il est très

¹⁹⁵ D. Borsboom, G. J. Mellenbergh, and J. van Heerden, "The Theoretical Status of Latent Variables," *Psychological Review* 110, no. 2 (2003).

probable de découvrir que la première a (collectivement) marqué beaucoup plus de paniers que la deuxième. Toutefois, dans ce contexte il est évident qu'on peut tirer aussi une conclusion ultérieure : si on prend deux joueurs, l'un très grand et l'autre petit, il est très probable de découvrir que le premier a marqué beaucoup plus de paniers que le deuxième. Bref, le même facteur explique les effets au niveau individuel *et* collectif : les grands marquent plus parce qu'il est vrai de chaque individu grand que c'est plus facile pour lui de marquer. D'autres exemples d'homogénéité locale pourraient être la corrélation entre la puissance musculaire et le poids soulevé, ou encore la taille et la hauteur en mètres d'une étagère qu'on arrive à atteindre. Qu'une population de gens très musclés soulève plus qu'une population d'individus grêles s'explique par le fait qu'un individu musclé normalement soulève plus qu'un individu grêle.

Dans la théorie traditionnelle des vertus et dans la théorie des traits, l'action causale des traits a été clairement conçue selon le modèle de l'homogénéité locale. Dans la littérature philosophique l'attention à des dynamiques psychologiques réelles qui intéressent l'individu est tout à fait explicite, car on se concentre sur le modèle de l'individu vertueux, qui est conçu comme un modèle prototypique à imiter. On ne part jamais d'une propriété intéressante qui émergerait à l'échelle d'une population (ex : il se trouve que parmi les militaires d'unités d'élites il y a moins de défections) pour se demander après quels sont les facteurs actifs au niveau individuel qui rendent cela possible (ex : quel ensemble de caractéristiques différentes – et pas forcément les mêmes pour tout sous-groupe de la population – permet l'émergence de l'effet global d'un bas taux de défection). Le sens de l'explication est toujours exactement le contraire : une vertu est d'abord quelque chose qu'un individu peut avoir ou ne pas avoir et les effets intéressants au niveau de la communauté sont ceux qui se produisent lorsque beaucoup de gens sont vertueux.

Même les théories de la vertu qui ne détaillent pas une psychologie morale informative et définissent la vertu comme « une bonne habitude » se réfèrent néanmoins sans doute au niveau individuel : il s'agit de bonnes habitudes qu'un individu peut posséder ou non, jamais de propriétés collectives et émergentes d'une communauté dont l'action à l'échelle individuelle reste mystérieuse ou non spécifiée¹⁹⁶. En tout cas, l'accent est mis sur l'action de la vertu dans le caractère vertueux, sur son interaction avec d'autres vertus et structures de personnalité – jamais une vertu n'est étudiée isolément et indépendamment des dynamiques psychologiques qu'elle détermine – et sur l'individu vertueux.

¹⁹⁶ Julia Driver, *Uneasy Virtue*, *Cambridge Studies in Philosophy* (Cambridge, UK ; New York: Cambridge University Press, 2001).

Un effet moyen qui émerge uniquement au niveau d'une population et n'a pas d'étiologie unique au plan individuel ne peut être étudié de cette façon : cela n'a pas de sens de se demander comment le courage interagit avec la justice dans le caractère vertueux, si on parle de propriétés de la population. En effet, il pourrait y avoir plusieurs formes de courage et de justice qui n'interagissent pas de la même manière chez des sous-groupes différents : dans chaque sous-groupe, les informations sur l'interaction spécifique entre vertus seraient perdues si on se limite à étudier l'interaction moyenne dans la population. De la même façon, il n'y a pas de modèle spécifique à imiter pour des propriétés localement non homogènes : dans une armée efficace uniquement grâce à l'organisation supérieure des détachements, il n'y a pas de soldat prototypique à imiter.

Je ne crois pas que les théoriciens de la vertu aient jamais nié l'existence de propriétés localement non homogènes sur l'échelle d'une population, ni que l'action d'individus vertueux soit impliquée dans leur émergence. Au contraire, ce dernier est un thème classique dans la liste des bénéfices attribués à la vertu : la diffusion d'une qualité comme le courage n'a pas uniquement l'avantage évident (et tautologique) de créer une population d'individus courageux, mais contribuerait aussi plus généralement à l'efficacité de la société. Cependant, dans les textes des théoriciens de la vertu, le courage lui-même n'est jamais associé au même statut émergent que cette forme générale d'efficacité : il s'agit clairement d'une attitude individuelle qui produit des effets fiables chez l'individu qui la possède, l'exemple classique étant le soldat ou le général.

L'approche classique en théorie des traits va dans le même sens, mais il est plus difficile d'étendre le présupposé d'homogénéité locale à tout le domaine, car la psychologie de la personnalité a traversé une crise des fondements que la théorie de la vertu n'a pas connue. Depuis les années 70, on cherche une façon adéquate à re-conceptualiser la personnalité et plusieurs propositions radicales et novatrices ont été mises sur la table. Il faut donc défendre l'idée selon laquelle le corpus des connaissances accumulées sur les traits présuppose l'homogénéité locale.

Si les propriétés qu'on retrouve en moyenne dans la population ne sont pas significatives au niveau individuel, par définition on n'a rien appris sur les processus à l'œuvre dans la psychologie réelle des individus, quelle que soit l'importance de l'effet global. Mais si le théoricien manifeste l'*intention* d'étudier des processus causalement actifs au niveau individuel, c'est qu'il doit présupposer l'homogénéité locale de la propriété étudiée. Démontrer un engagement réaliste (= processus actifs au niveau individuel) en ce sens est moins immédiat dans la littérature psychologique que dans la littérature philosophique (où l'engagement est explicite), mais, comme le

montre très bien Borsboom, la pratique standard des psychologues de la personnalité est incompréhensible et injustifiable si on n'est pas réaliste¹⁹⁷.

Pour commencer, dans le processus de validation psychométrique d'un test de personnalité¹⁹⁸, il est essentiel de montrer que des tests différents mesurent la même variable. S'il n'y a pas de tests différents qui donnent une image cohérente du niveau de névrotisme (ou de courage, ouverture d'esprit, etc.) des individus, c'est que le résultat du test de départ pourrait être un artefact lié à la structure des questions. Cette volonté de mesurer de plusieurs façons la même variable trahit une interprétation réaliste, car, par exemple, dans une interprétation vérificationniste radicale (opérationnisme), chaque test mesurerait des variables différentes¹⁹⁹.

Laissant de côté le vérificationnisme strict, position très peu plausible, on pourrait encore envisager des interprétations constructivistes, instrumentalistes ou fictionnelles des traits postulés par la recherche psychologique²⁰⁰. Mais là encore, à y regarder de près, la pratique des psychologues trahit un engagement réaliste, au sens où, sous l'hypothèse d'une ontologie différente, on n'arriverait pas à expliquer la méthodologie employée et l'absence de procédures alternatives. Si la préférence des théoriciens était allée à l'une des nombreuses nuances de non-réalisme sur les traits, ils auraient dû faire des choix méthodologiques différents : comme cela n'est pas le cas, il s'ensuit qu'on peut leur attribuer une position réaliste. Quels sont ces choix méthodologiques ? Sans trop entrer dans les détails techniques, la question principale est la suivante : l'adoption du modèle formel associé à la mesure de variables latentes telles que des traits de caractère n'est pas pleinement justifiée en dehors d'un cadre réaliste.

¹⁹⁷ Borsboom, Mellenbergh, and van Heerden, "The Theoretical Status of Latent Variables."

¹⁹⁸ C'est-à-dire le processus dans lequel on établit la possibilité de catégoriser les individus dans une certaine échelle évaluative où ils occuperaient une position plus au moins stable – ex : plus au moins névrotiques – et la capacité d'un questionnaire à estimer de façon fiable cette position.

¹⁹⁹ Dans une approche vérificationniste radicale (opérationnisme), une variable ou une entité théorique sont définies par la façon dont on peut les mesurer. Si l'idée paraît radicale dans les sciences physiques (la masse mesurée par la méthode *a* serait autre chose que la masse mesurée par la méthode *b*), elle est plus courante en sciences sociales. Par exemple, il n'est pas évident que le statut socio-économique mesuré d'une certaine façon soit la même « propriété » que le statut socio-économique mesuré d'une autre façon. Pour cette raison, en psychologie, la volonté de valider la même variable par plusieurs méthodes différentes trahit un engagement réaliste implicite. Voir U. Feest, "Operationism in Psychology: What the Debate Is About, What the Debate Should Be About," *Journal of the History of the Behavioral Sciences* 41, no. 2 (2005).

²⁰⁰ On retrouve le même problème dans l'interprétation réaliste ou anti-réaliste de la théorie des traits du sens commun, voir chapitre consacré à cette question.

Par modèle, nous entendons ici un système entièrement formel, non interprété. D'un point de vue formel, le modèle à variables latentes ne fait que construire une relation entre la valeur d'une variable et la valeur d'une série d'items (il peut s'agir du score de différentes questions dans un questionnaire). Une certaine distribution des scores des items est associée à une valeur de la variable latente : par exemple, les valeurs des réponses à n questions (dans une échelle de 0 à x , donc par exemple : 3,5,0,3,2) nous permettent de dériver la valeur correspondante de la variable latente.

Ces moyens formels nous permettent de faire deux choses : prédire l'ensemble des distributions probables des valeurs des items étant donné la valeur de la variable, ou estimer le paramètre (la valeur de la variable), étant donné la distribution des scores des items. Dans la pratique, cela correspond respectivement à : (1) prédire la distribution des réponses probables à un questionnaire sachant le positionnement dans l'échelle du trait que le questionnaire est censé mesurer ; (2) estimer le positionnement dans l'échelle du trait à partir des réponses au questionnaire, ce qui est d'ailleurs la raison d'être du questionnaire.

Pour l'instant, on se réfère aux tautologies non interprétées d'un système formel : les valeurs de la variable et des items sont liées par définition. Tout lien avec la réalité doit être justifié. Sans des présupposés réalistes, des ontologies constructivistes ou instrumentalistes n'arrivent pas à expliquer le choix de ce modèle formel. Le phénomène est assez général. Dans la pratique psychométrique, il y a fréquemment une pluralité et un surplus de modèles formels équivalents (ou presque) pour traiter le même corpus de données (qui donc ne permet pas de trancher entre les différents modèles) : le choix du théoricien est très révélateur quant à ses présupposés ontologiques²⁰¹. Par exemple, le genre de données qui ont justifié le succès du *big five*, le modèle d'analyse de la personnalité en cinq dimensions fondamentales qui s'est largement imposé à partir des années 90, auraient pu étayer un modèle avec six typologies de personnalité²⁰². La différence est remarquable, car une dimension de personnalité varie en degré (plus ou moins consciencieux) alors qu'on appartient ou pas à un type donné de personnalité.

²⁰¹ Ce n'est pas une application générale du célèbre principe de sous-détermination de la théorie par rapport aux données empiriques, selon lequel un nombre infini de théories différentes pourraient expliquer le même corpus de données et avoir ainsi la même adéquation empirique. Ici, le choix est réel, entre des modèles existants et praticables, alors qu'en philosophie des sciences il s'agit d'un choix hypothétique entre la théorie réelle et des théories possibles envisagées par le philosophe.

²⁰² P. C. M. Molenaar and A. von Eye, "On the Arbitrary Nature of Latent Variables," *Latent variable analysis* (1994).

Dans notre cas, le modèle des variables latentes (appelé aussi modèle réflexif²⁰³) est constamment préféré à d'autres modèles possibles, par exemple celui utilisé pour mesurer des propriétés émergentes comme la condition socio-économique. D'un point de vue formel, ce dernier modèle (appelé formatif) est exactement le contraire du modèle réflexif.

Intuitivement, cela signifie que le statut socio-économique change en fonction du changement de ses indicateurs, et pas l'inverse. C'est le fait d'avoir un meilleur travail et habiter dans un quartier plus sûr qui améliore le statut socio-économique. Ce n'est pas l'augmentation du statut socio-économique qui fait en sorte qu'on trouve un meilleur travail ou un quartier plus sûr où habiter. Dans le modèle formatif, la variable latente n'est pas conceptualisée comme quelque chose qui détermine la valeur probable des items mais comme une dimension qui *résume* d'une façon intéressante la mesure des items. Si les théoriciens avaient voulu conceptualiser l'ouverture d'esprit ou le caractère consciencieux de la même façon que le statut socio-économique, le choix du modèle formel pour traiter les données aurait été incompréhensible. Certes, nous n'avons pas encore exclu toute interprétation instrumentaliste ou constructiviste sophistiquée, mais il semble peu plausible d'en trouver une qui puisse être adéquate et attribuable de façon crédible à la plupart des théoriciens en psychologie de la personnalité. Par conséquent, nous sommes en mesure d'affirmer que – aussi bien en philosophie qu'en psychologie – les théories des traits *mainstream* et des vertus se sont développées dans le cadre d'une approche réaliste basée sur l'homogénéité locale des phénomènes étudiés.

Or, sachant que la presque totalité des données de la psychologie de la personnalité est constituée par des statistiques au niveau de groupes et populations²⁰⁴, il ne faut pas avoir des connaissances sophistiquées en statistique pour regarder avec suspicion le présupposé d'homogénéité locale. Car, on le sait bien, les tendances et les effets moyens effacent souvent des différences très intéressantes au niveau des sous-groupes de la population. Heureusement, on n'est pas laissé dans cet état de suspicion générale, car l'inadéquation de l'homogénéité locale a été suffisamment prouvée dans la littérature sur les traits, aussi bien d'un point de vue méthodologique qu'au niveau de nombreux exemples pratiques.

Avant de passer aux critiques méthodologiques, un exemple pratique peut aider. La timidité (*shyness*) est un trait général dont les sous-types ont été bien étudiés. Au fil des dernières décennies, on a proposé des distinctions entre (1) une timidité basée sur l'émotion de la peur (*fearful shyness*)

²⁰³ J. R. Edwards and R. P. Bagozzi, "On the Nature and Direction of Relationships between Constructs and Measures," *PSYCHOLOGICAL METHODS* 5, no. 2 (2000): 200.

²⁰⁴ Ce n'est pas impossible d'étudier vraiment et directement le niveau individuel, cela est – comme on le verra – seulement très difficile, et d'ailleurs on commence à le faire de plus en plus.

et une timidité provoquée par l'embarras et le fait d'être au centre de l'attention dans une situation sociale (*self-conscious shyness*)²⁰⁵ ; (2) une timidité qui amène à l'isolement social (*compulsive detachment* : renoncement à des relations et à des activités sociales) et une timidité caractérisée par le besoin de soutien et d'approbation de la part d'un nombre limité de personnes (*compulsive compliance*: suite à un investissement très sélectif dans un nombre limité de relations, l'individu est timide avec tous les autres. Il s'agit du type de personne qui est complètement perdue sans ses amis ou sa famille)²⁰⁶ et (3), pour absurde que cela puisse paraître, une timidité associée à la sociabilité (*shy-sociable*) et une timidité qui n'est pas associée à la sociabilité (*shy-unsociable*)²⁰⁷. Nous n'avons pas vocation à juger de la validité de ces sous-types ici, mais le dernier est spécialement pertinent par rapport à la question de l'homogénéité locale. Un timide sociable est quelqu'un qui aime et recherche les interactions sociales même s'il est fortement inhibé en société. Par exemple, il s'agit de quelqu'un qui va tout le temps à des soirées ou dans des boîtes de nuit, même si à chaque fois il n'interagit presque avec personne. Le timide pas sociable déteste les interactions sociales *parce qu'il est timide*. Dans ce contexte, il est évident que l'étude du « timide moyen » (des conséquences générales et moyennes de la timidité) a un intérêt très limité. Soit vous êtes un timide sociable, et vous chercherez constamment des interactions (comme un joueur de football « du dimanche » qui n'est pas du tout doué, mais qui poursuit avec détermination la gratification associée à la moindre réussite, fût-elle rare), soit vous êtes un timide pas sociable qui déteste les interactions. Le profil du « timide moyen » n'est ni très utile ni très informatif par rapport à la compréhension et à la prédiction du comportement d'un individu déterminé.

Des théoriciens tels que Molenaar ou Borsboom ont été récemment à l'origine de cette critique : ils soutiennent la thèse radicale selon laquelle la psychologie de la personnalité traditionnelle ne nous a *rien* appris sur la vraie psychologie des individus. Cette conclusion est exagérée : je vais montrer que, même s'il n'y a pas un seul genre de trait qui remplisse tous les rôles explicatifs envisagés par la tradition, il y a plusieurs genres de traits qui jouent des rôles plus limités, mais qui sont vraiment actifs dans la psychologie de l'individu. Dans ce cadre, la psychologie de la personnalité traditionnelle nous a surtout appris des choses intéressantes sur un type de trait spécifique : des tendances générales à l'action.

²⁰⁵ A. H. Buss, *Self-consciousness and social anxiety* (WH Freeman San Francisco, 1980).

²⁰⁶ J. M. Cheek et E. N. Krasnoperova, "Varieties of shyness in adolescence and adulthood," in *Extreme fear, shyness, and social phobia: Origins, biological mechanisms, and clinical outcomes*, édition L.A. Schmidt et J. Schulkin (New York: Oxford University Press, 1999).

²⁰⁷ J. M. Cheek, & Buss, A. H., "Shyness and sociability," *Journal of Personality and Social Psychology* 41 (1981).

En d'autres termes, en mobilisant la distinction de philosophie de la science entre réalisme quant aux entités (*entity realism* : la théorie postule des entités qui existent, mais les généralisations théoriques sur ces entités sont imparfaites, approximatives ou trompeuses) et réalisme quant à la théorie (les généralisations ou lois de la théorie sont vraies et décrivent la réalité)²⁰⁸ on peut dire que, par rapport aux traits (et contrairement à la thèse radicale des critiques) le réalisme quant aux entités est justifié (il y a bien des traits de ce genre) alors que les généralisations théoriques proposées ont été trop ambitieuses, imprécises et trompeuses à plusieurs égards.

Les traits généraux sont-ils des artefacts ? Défense d'un point de vue réaliste

Nous venons de montrer que la réflexion philosophique et psychologique sur les traits a adopté une approche réaliste. Cela relève d'une reconstruction correcte des intentions et des ambitions des théoriciens : il est temps d'évaluer le bien-fondé du projet. La question de l'homogénéité locale pose une menace redoutable. Faut-il pourtant en conclure que la théorie traditionnelle des traits n'a produit que des artefacts? C'est l'avis de critiques radicaux tels que Borsboom et Molenaar. Non contents de constater que la psychologie des traits traditionnelle n'étudie pas l'action des traits de façon adéquate (car elle s'intéresse uniquement à l'individu moyen et ignore les différentes interactions des traits dans les sous-groupes), ils nient purement et simplement l'existence de traits causalement actifs. Ils ont avancé trois critiques distinctes :

- (1) une attitude qui ne change pas dans le temps ne peut avoir des effets causaux, or, le niveau des traits est relativement stable par définition, donc, les traits n'ont pas d'effets causaux ;
- (2) s'il existe des traits, alors toute variation dans leur manifestation devrait être expliquée par une variation correspondante du niveau du trait, mais, dans les questionnaires, les gens ne scorent pas moins (disons) en timidité les jours où ils se comportent de façon moins timide : donc les traits n'existent pas ;
- (3) si les traits existent, alors la psychologie des personnes qui sont caractérisées par un certain niveau d'un trait devrait être significativement semblable (ex : le fonctionnement psychologique de deux personnes moyennement timides devrait être presque identique), ce qui n'est pas le cas : donc les traits n'existent pas.

(1) Le premier argument est le suivant : un facteur ne peut être causalement actif chez un individu que s'il explique la variation interindividuelle réelle ou hypothétique, c'est-à-dire la variation de

²⁰⁸ N. Cartwright, *How the Laws of Physics Lie* (Oxford University Press, 1983).

performance chez le même individu d'un moment à l'autre. Dans cette formulation de leur critique, ils s'intéressent plutôt à la variation réelle et ils semblent soutenir que – comme la plupart des conceptions de la causalité postulent une *covariation* de la cause et de l'effet – s'il n'y a pas de variation, il ne peut pas y avoir de causalité :

« La position d'un individu dans une échelle qui mesure la valeur d'une variable latente est, dans le modèle standard de mesure, conceptualisée comme une constante, et une constante ne peut pas être une cause²⁰⁹ ».

Par construction de la notion de trait, un individu a un niveau plus ou moins stable de timidité, donc la timidité ne peut expliquer causalement le comportement. Par contre, si on considère l'humeur – qui est un *état* psychologique et non un trait – on voit qu'elle change dans le temps, et ses effets la suivent. On peut dire que l'humeur cause ses manifestations chez l'individu.

Or, cette référence à la variation réelle est clairement trop stricte, car, par exemple, la taille peut être responsable du fait qu'on atteigne les livres sur une étagère, et cela même si la taille d'un individu adulte ne change pas plus rapidement que ne le fait sa personnalité. C'est la variation hypothétique ou contrefactuelle qui compte. Ainsi, il est clair que « si j'avais été plus grand, j'aurais pu atteindre une étagère plus élevée », « si j'avais été une personne plus névrotique, j'aurais crié », etc. Ces exemples montrent le pouvoir explicatif de variables dont le niveau réel est stable. En ce qui concerne la variation réelle, les fluctuations dans la manifestation du trait ne sont pas pertinentes : elles ne peuvent être que du bruit par rapport à la stabilité de la variable latente. L'existence de micro-fluctuations hebdomadaires de mon niveau de timidité n'est pas significative.

Borsboom admet que la causalité est compatible avec un manque de variabilité réelle, mais ensuite – de façon assez incompréhensible – dans la deuxième formulation de sa critique, il semble soutenir la thèse suivante : s'il y a causalité, alors n'importe quelle forme de variation *réelle* doit être expliquée par une variation de la variable.

(2) A propos de l'intelligence (qui est assimilée ici à un trait et donc ne devrait pas avoir de pouvoir causal, d'après Borsboom), du test du QI et du contrefactuel « Si Einstein avait été moins intelligent, il aurait donné des réponses différentes au test du QI », Borsboom affirme :

« On peut imaginer que si le modèle intra-sujet pour Einstein avait la même structure que le modèle inter-sujets, alors la thèse de la causalité individuelle aurait du sens malgré les difficultés qu'on a rencontrées ci-dessus.

²⁰⁹ Borsboom, Mellenbergh, and van Heerden, "The Theoretical Status of Latent Variables," 211.

Essentiellement, cette situation aurait pour effet que la façon dont Einstein est différent de lui-même au fil du temps et la façon dont il est différent des autres à un moment donné soient les mêmes²¹⁰. »

L'idée est la suivante : la notion d'intelligence est adéquate pour rendre compte des différences entre échantillons ou populations (les scientifiques sont plus intelligents que les mécaniciens), mais, pour qu'on puisse soutenir que l'intelligence est active au niveau individuel, il faudrait que les différences dans les prestations cognitives d'un même sujet d'un jour à l'autre soient expliquées par des fluctuations de l'intelligence. Dans le cas de la personnalité : si l'extraversion ou le névrosisme sont actifs chez l'individu, il faudrait que des variations dans les manifestations de ces traits d'un jour à l'autre chez le même individu soient dues à une fluctuation du niveau d'extraversion ou névrosisme.

Nous ne voyons pas la pertinence du raisonnement. Par définition, on sait déjà que le niveau des traits est stable sur le court terme et que des variations d'un jour à l'autre doivent être attribuées à d'autres facteurs, tels que la fatigue ou le stress. L'idée selon laquelle l'intelligence ou la personnalité changent d'un jour à l'autre n'a simplement pas de sens. Ce serait comme soutenir la thèse suivante : si la taille explique la différence entre les hauteurs moyennes auxquelles deux populations peuvent sauter, alors (pour que la taille soit un facteur explicatif véritable à l'échelle de l'individu) il faudrait que les différences entre plusieurs sauts de la même personne à des moments différents de la journée soient expliquées par un changement de taille... Suivant cette logique, Borsboom et Molenaar ont conduit des expériences pour montrer que la variation temporelle intra-individuelle ne suit pas les axes du *big five* ou des facteurs de l'intelligence générale : tout cela dans le but de discréditer l'approche traditionnelle des traits²¹¹. Mais le rôle des traits n'est pas celui d'expliquer les fluctuations temporelles.

Pour l'instant, il n'y a pas d'argument convaincant pour nier que le contrefactuel « Si Einstein avait été moins intelligent, il aurait donné des réponses différentes au test du QI » puisse exprimer un lien causal.

(3) Il existe une troisième version de la critique. Borsboom semble soutenir qu'un trait pourrait être causalement actif si la « réalisation » dans la psychologie individuelle d'un certain niveau du trait était la même pour tout le monde : en d'autres termes, si les processus psychologiques de toutes les

²¹⁰ Ibid.: 212-213.

²¹¹ P. Borkenau and F. Ostendorf, "The Big Five as States: How Useful Is the Five-Factor Model to Describe Intraindividual Variations over Time?," *Journal of Research in Personality* 32, no. 2 (1998).

personnes qui sont moyennement intelligentes (disons $QI\ x$) étaient exactement les mêmes. Si cela était ainsi, on pourrait considérer la notion d'intelligence comme un raccourci commode pour se référer aux processus qui sont actifs au niveau individuel chez tous les gens qui ont le même niveau d'intelligence (processus qui devraient être identiques). Egalement, en ce qui concerne la personnalité, cela signifierait qu'exactly les mêmes dynamiques psychologiques seraient à l'œuvre chez les individus qui ont le même niveau d'extraversion. Les faits ne confirment pas cette prédiction :

« Jane, qui a un niveau élevé dans la dimension d'intelligence générale inter-sujets, va très probablement approcher les questions du test du QI avec une stratégie qualitativement différente par rapport à son frère John. John, et son neveu Peter, également intelligents, peuvent répondre de façon incorrecte à la même question, mais pour des raisons différentes (John a des problèmes à se souvenir des séries de patterns dans le test de Raven, alors que Peter a difficulté à imaginer des rotations spatiales). Il est évident que le problème est encore plus sérieux pour les tests de personnalité, dans lesquels en général on n'a pas la moindre idée de ce qui se passe entre la lecture de la demande et la réponse²¹². »

Si on pouvait regarder ce qui se passe dans la tête d'Einstein lorsqu'il répond à un questionnaire d'intelligence – insiste Borsboom – on verrait certainement l'activation de la mémoire à court et long terme, de la capacité de calcul ou de l'imagination spatiale :

« mais où l'intelligence entre-t-elle dans le processus qui amène à la réponse de Einstein ? La réponse est nulle part²¹³. »

Cet argument prouve trop. En un sens, il est correct. Les notions « méta », qui ne se réfèrent qu'à une série des processus de plus bas niveau, semblent ne rien ajouter à l'ontologie du monde. En tout cas, rien qu'on ne retrouve déjà directement chez les processus de bas niveau eux-mêmes. Si on regarde à la loupe, à un niveau d'analyse suffisamment fin, il n'y a que l'activation de processus et facultés cognitives très spécifiques : personne ne le nie. Mais il faut opérer une distinction entre le cas où les processus de bas niveau n'ont presque aucun lien entre eux et le cas où ils co-occurrent de manière à la fois systématique et significative. Imaginons que quelqu'un cherche des amis qui aient au moins deux ou trois de ces qualités : fêtard, très cultivé, à l'écoute, bon sens de l'humour, fiable, disponible, un peu fou mais sans excès, etc. Quelqu'un qui a deux ou trois de ces qualités de façon très marquée sera considéré comme étant « bien » ou « cool ». Est-ce qu'une « attitude cool »

²¹² Borsboom, Mellenbergh, and van Heerden, "The Theoretical Status of Latent Variables," 214.

²¹³ Ibid.: 213. La question de l'interprétation de l'intelligence (le QI, le facteur g) a fait l'objet d'un débat spécifique et trop complexe pour que nous puissions l'aborder ici. L'usage de l'exemple de l'intelligence est justifié parce que l'intelligence est mesurée par les mêmes méthodes que les traits de personnalité généraux.

en ce sens est une propriété causalement active dans la psychologie des gens ? On peut en douter, car ce concept de cool n'est rien d'autre qu'un assemblage arbitraire d'attitudes réelles qui n'ont pas beaucoup de liens entre elles. A chaque fois, quelqu'un de cool sera soit « fêtard et cultivé, mais avec un mauvais sens de l'humour et peu fiable » soit « disponible et capable d'écouter, mais timide et peu cultivé », etc. Le concept de « cool » n'ajoute rien et ne fait aucun travail explicatif : ici la suspicion par rapport au passage au niveau méta est justifiée. Mais considérons le processus de vieillissement d'une personne, le processus global qui amène inexorablement à la mort. D'un point de vue biochimique et médical, il n'y a pas *le* vieillissement, mais une série de processus différents selon qu'il s'agisse de tissus et organes différents, ainsi que des causes biochimiques différentes (accumulation de dommage dans l'ADN nucléaire ou mitochondrial, accumulation de micro-débris dans le milieu intra-cellulaire ou extra-cellulaire, etc.). Un homme peut mourir à la soixantaine avec les artères et les poumons d'un octogénaire, alors qu'un ultra centenaire peut mourir de simple arrêt cardiaque avec des reins et des artères de quelqu'un sur la soixantaine.

Est-ce que cela nous autorise à penser qu'il n'y a pas de processus de vieillissement global causalement actif au niveau de l'individu ? Non, la différence est la suivante : les processus de bas niveau co-occurrent de façon suffisamment systématique et significative. Lorsque cela arrive, dans l'usage commun, dans la pratique scientifique et en philosophie des sciences²¹⁴, l'existence et la force causale du processus de haut niveau sont communément admises. D'ailleurs, le rejet de cette approche produit une régression paradoxale, car tout processus n'est qu'une façon de résumer à un méta-niveau l'action de facteur de plus bas niveau, et ainsi de suite. Dans l'exemple, Borsboom considère la mémoire à long terme comme un cas paradigmatique de mécanisme réel dans la psychologie de l'individu, mais sommes-nous sûrs que récupérer un souvenir mobilise les mêmes stratégies cognitives de très bas niveau chez tous les individus ? Dans le vocabulaire pointu des sciences cognitives, « se souvenir » n'est qu'une expression qui résume de façon grossière une vaste panoplie de stratégies de récupération.

Cette thèse n'a pas vocation à défendre dans les détails une philosophie de la psychologie déterminée (travail qui demanderait tout l'espace à disposition) et doit se limiter à évaluer le statut de la notion de trait par rapport à un arrière-plan de positions méthodologiques et philosophiques communément admises²¹⁵ : pourtant, je ne défendrai pas davantage l'interprétation réaliste

²¹⁴ C'est la position que les réalistes de Cornell (Boyd, Sturgeon, Brink) appliquent à la science et à la morale, voir R. Boyd, "How to Be a Moral Realist," *Essays on Moral Realism* (1988).

²¹⁵ En d'autres termes, on ne s'occupe que des problèmes spécifiques à la notion de trait, alors que des positions radicales sur la causalité et les explications psychologiques créent des problèmes pour les traits *et* pour beaucoup

(=causale) d'un système homéostatique de facteurs co-occurents. Par contre, il faut montrer que le modèle s'applique aux traits de personnalité.

Or, il me semble que, y compris dans le cas de super-traits très généraux comme les dimensions du *big five*, la co-occurrence des facteurs de plus bas niveau qui figurent dans la définition des traits est suffisamment forte et significative pour parler d'existence réelle. Le cas reste débattu parce que, bien sûr, il n'y a pas un réseau de co-occurrences aussi dense et inéluctable que celui du vieillissement. Cependant, ces corrélations sont plutôt stables ou se développent de manière cohérente au fil de la vie, et elles sont significatives. Bien sûr, il est souvent plus opportun de mobiliser des notions plus spécifiques que les tendances générales étudiées par la psychologie de la personnalité classique. Trop d'ambitions pèsent sur les dimensions générales du *big-five*. Mais ici nous sommes uniquement en train d'évaluer la thèse réaliste selon laquelle les individus *ont* vraiment ces tendances générales, aussi modeste soit leur importance.

Il semble raisonnable d'admettre que ces tendances au comportement sont associées à des effets et conséquences spécifiques. À long terme, une tendance générale à la timidité peut avoir des conséquences spécifiques qui ne se réduisent pas aux effets de la configuration particulière de timidité (timide avec des femmes ou avec des hommes, avec les supérieurs au travail ou avec une nouvelle connaissance, etc.) de l'individu concerné. Par exemple, savoir qu'une personne a une tendance générale à la timidité (en plus de la liste précise de situations dans lesquelles elle est timide) peut aider à faire des prédictions sur son évolution et ses réactions futures. De la même façon, savoir qu'une voiture a tendance à développer certains problèmes est mieux qu'avoir uniquement une liste des routes en bonnes conditions où on peut conduire sans soucis.

Cela est d'autant plus vrai d'un point de vue évolutif : il est certain que l'évolution favorise la spécificité, mais, comme les défenseurs du *big five* le soutiennent, dans des cas spécifiques (variabilité environnementale permanente) elle peut favoriser aussi des tendances générales. En d'autres termes, une tendance générale à la timidité pourrait être une adaptation sélectionnée parce qu'elle favorise la *fitness* dans certains environnements sociaux : cela est compatible avec la nature adaptative (pour des raisons différentes) des mécanismes plus spécifiques qui sont responsable des variations situationnelles dans le niveau de timidité²¹⁶.

d'autres notions psychologiques. On s'occupe uniquement du caractère problématique *différentiel* de la notion de trait.

²¹⁶ Certains psychologues évolutionnistes pensent que la sélection sexuelle (la sélection qui ne favorise pas des traits qui aident un individu à survivre plus que les autres individus, mais des traits qui l'aident à avoir plus de succès dans la compétition sexuelle) peut favoriser l'émergence de traits généraux. Un trait général comme l'extraversion serait un indice de « bonne qualité génétique » plus fiable qu'une capacité spécifique à gérer des tensions au sein d'une

Au stade actuel, ce ne sont là que des hypothèses, mais il me semble prématuré de considérer les tendances générales des individus postulées par la psychologie de la personnalité comme de purs artefacts. Il est certain, par contre, que la psychologie n'a été capable d'étudier ces traits généraux que dans l'individu moyen, ce qui n'aide pas les ambitions de la théorie de la vertu.

Les effets holistes sont encore moins fiables que ceux associés à un seul trait

Vérifier que l'action d'un facteur est isomorphe aux niveaux de la population et de l'individu resterait un problème pressant même si on était en train d'analyser une variable unidimensionnelle, qui influence le comportement indépendamment d'autres facteurs : l'effet de cette variable pourrait être différent selon les sous-groupes de la population. Mais les traits, aussi bien en philosophie qu'en psychologie, sont des phénomènes psychologiques actifs dans un contexte fortement holiste : leurs effets sont additifs et ils interagissent constamment.

Même dans le cas le plus simple, les chercheurs ont toujours insisté sur la nature additive de l'action des traits : on n'est pas courageux ou honnête avec la même intensité et détermination suivant qu'on est aussi timide ou extraverti. Plus couramment, c'est une véritable interactivité qui est proposée : l'individu courageux qui n'est pas aussi juste ne manque pas de peu la marque de la vertu, mais il se montre totalement incapable d'agir de façon correcte. Dans certains contextes sensibles, l'individu courageux mais pas juste ne peut pas exhiber la conduite requise par le courage : par exemple, il peut attaquer un ennemi qui mérite d'être épargné²¹⁷.

Le holisme de l'action des traits et des vertus démultiplie l'écart entre l'étude de l'individu moyen et la compréhension des dynamiques à l'œuvre dans la psychologie réelle des gens. Face à une seule dimension, disons la timidité, on peut se demander si la description moyenne du trait n'occulte pas des variantes intéressantes : par exemple, deux ou trois façons d'être timide non équivalentes. En un sens, le problème paraît modeste et on pourrait s'attendre uniquement à des variations d'intensité dans la manifestation de la phénoménologie du trait. Mais, en ce qui concerne l'économie d'interaction complexe de plusieurs traits dans une personnalité, l'abstraction à l'individu moyen peut être responsable d'une série d'artefacts. Il est possible que certains schémas d'interaction – la générosité a besoin de l'extraversion pour se manifester correctement dans des contextes sensibles –

relation d'amitié. Nos ancêtres ont peut-être eu un intérêt spécifique à remarquer les tendances générales et peu fiables des partenaires. C'est l'approche de Miller à l'évolution des traits généraux, y compris l'intelligence. Voir G. F. Miller, *The mating mind: How sexual choice shaped the evolution of human nature* (Doubleday Books, 2000).

²¹⁷ La question est largement analysée dans le cadre du débat sur l'unité de la vertu, voir chapitre sur la sagesse.

soient un artefact de la moyenne et disparaissent complètement chez les individus réels. Il n'y aurait pas là une variation de degré ou d'intensité, mais une erreur catégorielle : la postulation d'un type d'interaction qui n'a pas lieu dans la réalité. Celle-ci est une preuve ultérieure et beaucoup plus nette de l'inadéquation de la théorie des traits généraux, du point de vue des ambitions de la théorie de la vertu. Dans un modèle du courage et de la justice conçus comme des traits généraux, les interactions découvertes par l'étude expérimentale pourraient ne pas exister à l'échelle de l'individu.

Approches psychologiques de l'étude de la vertu

Malheureusement, à l'heure actuelle, les seules approches systématiques existantes pour l'étude de la vertu adoptent le modèle des traits généraux, héritant de toutes les difficultés qu'on vient de décrire. La recherche empirique contemporaine sur le fonctionnement des traits de caractère considérés comme étant « vertueux » peut être divisée en deux ensembles. D'un côté, il y a une étude – méthodologiquement très hétéroclite – de chaque « vertu » prise séparément. Ainsi, les recherches sur les bases cognitives, génétiques, et sur la structure psychologique de la créativité, la curiosité, l'ouverture d'esprit, la compétence sociale (etc.) se multiplient. Dans chacun de ces exemples, la vertu concernée est définie et étudiée de plusieurs façons par des chercheurs différents. Par exemple, toute manifestation de comportement ou d'attitudes intellectuelles qui peut être associée à la créativité, en un sens très large, est prise en compte. Cette recherche est l'équivalent contemporain du débat sur chaque vertu particulière (courage, justice) qu'on retrouve dans la tradition philosophique, indépendamment de l'allégeance à des modèles généraux du fonctionnement du caractère.

De l'autre côté, des psychologues ambitieux ont abordé directement la question de la *structure* du caractère vertueux. Cette deuxième forme de recherche est très minoritaire. À notre connaissance, il y a deux exemples à citer : la classification VIA développée par Seligman et Peterson, et l'étude lexicale des termes des traits vertueux par Cawley.

Comment expliquer l'état actuel de la recherche empirique ? (1) La psychologie empirique s'est toujours intéressée beaucoup plus au dysfonctionnement pathologique qu'au fonctionnement optimal de la cognition humaine. Dans les années 90, une réaction à cette situation a donné naissance au mouvement très en vogue de la « psychologie positive », qui inclut l'étude des émotions positives, de la satisfaction personnelle, du bonheur et aussi des vertus. (2) Même à l'intérieur de ce cadre, les vertus occupent une place plutôt problématique, à cause de leur caractère évaluatif très marqué et du désaccord sur la liste des vertus. C'était pour cette raison que les termes

de traits qui pouvaient être associés à des vertus avaient été écartés par Allport, le fondateur du domaine de la psychologie de la personnalité.

Dans la tradition du *big five*, les super-traits ou macro-dimensions de la personnalité sont dérivés par analyse factorielle à partir d'une liste de termes de traits tirés d'un dictionnaire (approche lexicale). Depuis le début, faute d'accord sur une liste, les termes de traits vertueux ont été délibérément exclus de l'analyse factorielle : la psychologie pensait établir sa crédibilité grâce à une division nette entre le domaine neutre de la personnalité et celui controversé du caractère moral. Les séquelles de cette attitude sont évidentes aujourd'hui : en ce qui concerne l'étude des « vertus » prises séparément, on garde un profil bas en parlant de « traits positifs », sans recourir au terme technique de vertu, avec son bagage théorique et évaluatif. *A fortiori*, on ne s'aventure presque jamais dans une modélisation directe du fonctionnement des vertus.

La distinction nette entre personnalité et caractère moral peut se révéler dangereuse pour la psychologie de la personnalité. Dans le cadre de l'approche des traits, un désaccord sur la valeur morale n'est pas une raison suffisante pour exclure une dimension fondamentale de la personnalité. Nous rappelons que les dimensions du *big five* définissent des *clusters* de phénomènes psychologiques qui tendanciellement co-occurrent. Chez quelqu'un qui a un niveau élevé d'extraversion, une série de manifestations comportementales et émotionnelles ont tendance à se produire. Proposer un modèle à n dimensions revient à proposer un découpage optimal des manifestations significatives de la personnalité en n sous-ensembles, tels que les éléments de chaque sous-ensemble sont associés plus souvent entre eux que ne le sont avec les éléments des autres sous-ensembles. Par exemple, le fait qu'un modèle en cinq facteurs soit meilleur qu'un modèle à quatre facteurs signifie qu'un découpage en cinq maximise en même temps la « cohésion » dans chaque sous-ensemble et la « divergence » entre sous-ensembles différents. Or, avant d'effectuer une analyse factorielle, il est important de ne pas sur-représenter certains termes de trait. En ce sens, les psychologues ont raison de craindre des biais moraux. Si, dans un système moral qui met l'accent sur le courage et la loyauté, on sur-représente les termes proches de ces notions, le résultat en sera affecté.

Mais exclure tous les termes évaluatifs n'est pas une solution. Quelle que soit la valeur morale du courage, les manifestations comportementales associées à cette vertu font partie de la personnalité et doivent être prises en compte dans l'analyse factorielle. Notre indécision dans l'évaluation morale d'une dimension fondamentale de variation individuelle n'est pas un argument recevable contre son existence. Par exemple, le courage pourrait être une macro-dimension de variation individuelle, quel que soit notre niveau d'appréciation. Autrement dit, les psychologues des traits avaient de bonnes raisons d'ignorer le statut non consensuel de certains termes de traits, mais pas

d'ignorer leur existence. Pour donner un autre exemple, imaginons qu'une culture considère la timidité comme immorale : d'un point de vue purement descriptif, ce n'est pas pour cela que la timidité cesserait d'être un trait de personnalité.

C'est là un argument contre la complétude du *big five*. Si les termes dénotant des vertus avaient été inclus dans l'analyse factorielle, on aurait peut-être trouvé plus de dimensions, ou des dimensions différentes. Il reste une seule possibilité pour défendre le *big five* : si les termes de traits vertueux forment une classe à part, on a le droit de les étudier séparément et de ne pas les mélanger avec les termes de traits « normaux ». Selon cette hypothèse, les termes de trait « fortement évaluatifs » formeraient une « espèce naturelle psychologique » distincte des traits normaux. Cette approche peut nous amener à deux résultats : (1) la découverte que, du point de vue descriptif, les vertus peuvent se réduire à une certaine distribution de traits, ce qui justifie une réduction des vertus aux traits de personnalité ou (2) une classification des vertus qui serait complètement autonome par rapport à la classification des traits de personnalité.

Le premier cas de figure est celui envisagé par la psychologie de la personnalité classique. D'après Allport, le caractère n'est que la « personnalité évaluée »²¹⁸. Par-delà l'évaluation, il n'y aurait aucun élément descriptif nouveau introduit par le discours sur les traits. Toute vertu serait l'ensemble de plusieurs traits de personnalité, plus une évaluation. Par exemple : une personne courageuse n'est qu'une personne extrovertie et déterminée. Ce mélange spécifique d'extraversion et de détermination fait l'objet d'une approbation morale. En ce sens, l'étude des vertus n'ajoutera rien de significatif à une compréhension psychologique de la personnalité.

Dans le deuxième cas de figure (une classification des vertus *indépendante* de la classification des traits de personnalité), on aboutit à une réforme du modèle en cinq facteurs qui ajoute n facteurs de vertu, tout en préservant l'identité des facteurs de personnalité. Cela donnerait un modèle en « 5 (personnalité) + n (vertu) » facteurs. Si par « personnalité » on veut se référer à la dimension de variation psychologique individuelle dans sa globalité, il n'est plus vrai qu'il y a cinq facteurs, mais cinq (faiblement évaluatifs) plus n (fortement évaluatifs) facteurs. Néanmoins, ce modèle garderait une certaine indépendance pour la personnalité « non vertueuse » : les résultats du *big five* seraient corrects mais partiels. En revanche, si on devait mélanger tous les termes de traits (faiblement et fortement évaluatifs) avant l'analyse factorielle, la structure en cinq facteurs risquerait de disparaître complètement.

²¹⁸ Allport, *Personality; a Psychological Interpretation* 52.

Les travaux des Cawley s'inscrivent dans ce contexte²¹⁹. Il a cherché dans le dictionnaire des termes de traits vertueux qui avaient été exclus par les approches lexicales traditionnelles. Les termes ont été sélectionnés par le critère suivant : si l'insertion d'un adjectif dans l'expression « sois ... » donnait une recommandation normative crédible (sois patient, sois courageux), l'adjectif pouvait être considéré comme l'expression d'une vertu. Le résultat de ce travail est une liste de 140 termes qui a été transformée dans un test où chaque terme correspond à un item représenté par trois phrases. Par exemple, pour « prudent » :

« Je suis prudent dans mes affaires quotidiennes. Je suis capable de prendre des décisions correctes à partir de mon jugement et du sens commun. Les personnes qui me connaissent le plus me décriraient comme étant prudent, discret et raisonnable²²⁰ ».

Pour chaque item, les participants devaient s'auto-évaluer sur une échelle de un à sept. Suite à une analyse factorielle, quatre macro-dimensions propres au domaine de la vertu ont émergé. L'empathie, l'ordre, le fait d'être plein de ressources (*resourceful*) et la sérénité. À titre d'exemple, à l'empathie correspondent les termes suivants : amical, sensible, compassionnel, courtois, etc. À l'ordre : austère, sérieux, discipliné, décent, etc. Au fait d'être plein de ressources : persévérant, sagace, intelligent, confiant, etc. À la sérénité : serein, pacifique, patient, etc.

Le domaine des traits fortement évaluatifs aurait une structure en quatre facteurs qui ne sont pas réductibles au *big five*, car ils ne sont pas étroitement corrélés à ses cinq dimensions, comme cela devrait être le cas si une réduction était possible²²¹. En ce sens, l'hypothèse classique selon laquelle le caractère ne serait que la « personnalité évaluée » paraît écartée. Il s'agirait d'accepter un modèle en neuf facteurs : cinq dimensions de personnalité et quatre de vertu.

Cette proposition est intéressante, mais il nous semble que Cawley n'a pas suffisamment justifié la nécessité de séparer en deux classes les termes de trait « faiblement » et « fortement » évaluatifs. Selon le sens commun, tous les traits sont au moins faiblement évaluatifs, et l'intensité de cette évaluation semble créer un continuum plutôt que deux classes étanches. On ne voit pas de raisons crédibles pour ne pas mélanger tous les traits avant l'analyse factorielle. Cette procédure qui, à notre connaissance, n'a jamais été expérimentée, pourrait mettre sérieusement en péril le modèle en cinq facteurs.

²¹⁹ M. J. Cawley, J. E. Martin, and J. A. Johnson, "A Virtues Approach to Personality," *Personality and Individual Differences* 28, no. 5 (2000).

²²⁰ *Ibid.*: 1002.

²²¹ *Ibid.*: 1006.

De toute façon, l'approche de Cawley hérite de tous les problèmes de l'approche traditionnelle des traits, notamment la non-fiabilité des traits généraux, le manque d'homogénéité locale et l'impossibilité d'étudier l'interaction entre vertus chez l'individu.

La classification VIA

Le deuxième exemple de modélisation empirique de la vertu représente un effort considérable qui mérite d'être salué. Les psychologues « positifs » Peterson et Seligman ne cachent pas leurs ambitions : ils veulent établir, pour l'excellence du fonctionnement de l'esprit humain, l'équivalent de la classification DSM des troubles psychiques. Le DSM est un outil de classification reconnu au niveau international ; il a été développé pour disposer d'un schéma de classification qui puisse être appliqué et compris partout dans le monde. Les développeurs de la classification des vertus VIA (*value in action*) prennent l'analogie très au sérieux. Il s'agit d'une liste de 24 vertus que les auteurs cherchent systématiquement à valider et à imposer comme le nouvel étalon dans le panorama des nouvelles pratiques de mesure des « attitudes positives ».

Au fil des dernières années, dans le secteur des sciences humaines appliquées, on s'est rendu compte que le fonctionnement optimal de la psychologie humaine est beaucoup moins bien compris, cartographié et mesuré que la dimension pathologique, d'où la nécessité très pratique d'outils psychométriques plus raffinés pour mesurer la satisfaction, le bonheur, les émotions positives et d'autres facteurs essentiels au succès d'entreprises et d'associations. Dans ce nouveau marché, la classification VIA semble être destinée à jouer un rôle important. Ses défenseurs sont en train d'étudier les effets, les corrélats, les influences génétiques et environnementales, la variation culturelle et la dimension développementale des 24 vertus²²². Ce travail de validation impressionnant représente l'effort empirique le plus sérieux qui ait jamais été fait pour étudier la vertu, et il mérite d'être analysé avec attention.

Malheureusement, ces recherches se placent explicitement dans le cadre théorique de l'approche des traits : rien n'est fait pour résoudre les problèmes analysés dans les sections précédentes. Seligman et Peterson n'ont pas adopté la méthode lexicale : leur liste des vertus n'est pas issue d'une étude des termes de trait dans les langues naturelles, mais plutôt d'un vaste effort de recherche de vertus universelles, dans plusieurs cultures et sous-cultures. Ainsi, une équipe de

²²² K. Otake et al., "Development of the Japanese Version of the Values in Action Inventory of Strengths (Via-Is)," *Shinrigaku Kenkyu* 76, no. 5 (2005), N. Park and C. Peterson, "Moral Competence and Character Strengths among Adolescents: The Development and Validation of the Values in Action Inventory of Strengths for Youth," *Journal of Adolescence* (2006), M. F. Steger et al., "Genetic and Environmental Influences on the Positive Traits of the Values in Action Classification: Findings from a Twin Study," *Journal of Research in Personality* 41 (2007).

psychologues a étudié la littérature anthropologique, philosophique, historique et sociologique pour découvrir un noyau de vertus universelles. L'attention n'a pas été réservée à quelques cultures représentatives : par exemple, ils ont étudié aussi bien le code de chevalerie de Charlemagne que le code éthique des Klingons de Star Trek.

Malgré le rejet d'une stratégie lexicale, leur allégeance au modèle théorique du *big five* est totale : « Notre approche au caractère est dans l'esprit de la psychologie de la personnalité, en particulier la théorie des traits, mais non la caricature mise en avant et après critiquée par les théoriciens de l'apprentissage sociale dans les années 70. Au contraire, nous nous basons sur la nouvelle psychologie des traits, qui reconnaît l'existence de différences individuelles générales et stables, mais aussi influencées par l'histoire de l'individu et ainsi capables de changer²²³ ».

La référence à la synthèse contemporaine en théorie des traits est claire. Le modèle du *big five* se retrouve presque au pied de la lettre dans la structure des vertus VIA. En psychologie de la personnalité, il s'agit d'organiser sur une série de niveaux (et de façon récursive) les différents éléments qui covarient dans le fonctionnement psychologique de l'individu, à partir des attitudes et des comportements les plus spécifiques jusqu'aux tendances les plus générales. Un trait *x*, à un certain niveau d'abstraction, représente la covariation d'un ensemble de traits de niveau inférieur, alors qu'un trait *y* de niveau supérieur exprime la covariation de *x* et d'autres traits du même niveau. En d'autres termes, on part de réactions émotionnelles ou comportementales déterminées, par exemple se fâcher avec son patron lorsqu'on reçoit une critique (réponse spécifique) et on passe à un ensemble de ces réponses, qui forment une réponse habituelle, par exemple, être agressif au travail. Un groupe de réponses habituelles qui ont tendance à se produire ensemble forme une facette ou trait de premier niveau (ex : agressivité). Un cluster de facettes forme un super trait, ce qui correspond à une des dimensions du *big five* (ex : extraversion).

Dans la classification VIA on retrouve les trois niveaux, définis par la même procédure. Les thèmes situationnels (*situational themes*) correspondent aux réponses habituelles et sont relatifs à un certain contexte. Peterson et Seligman citent l'exemple du lieu de travail : anticiper les besoins des autres, faire en sorte que les gens se sentent intégrés dans le groupe, voir ce qu'il y a de positif dans le contexte de travail et chez ses collaborateurs. Les 24 clusters de thèmes situationnels correspondent aux facettes du *big five*. Pour posséder un de ces traits de premier niveau, il faut afficher un bon nombre de thèmes situationnels correspondants. Pour la vertu de la curiosité, il faut être curieux dans un nombre suffisant de contextes. Les six *master virtues* correspondent aux cinq super facteurs

²²³ Christopher Peterson and Martin E. P. Seligman, *Character Strengths and Virtues: A Handbook and Classification* (Washington, DC New York: American Psychological Association ; Oxford University Press, 2004) 10.

du *big five*. Pour avoir un niveau significatif dans un ou plusieurs de ces super traits, il faut posséder plusieurs des facettes correspondantes.

La liste des vertus est la suivante : Sagesse et connaissance (créativité, curiosité, jugement critique, envie d'apprendre, mise en perspective), courage (courage, persistance, authenticité, vigueur physique et intellectuelle), humanité (gentillesse, capacité à aimer, intelligence sociale), justice (capacité de travailler en groupe, justice, leadership), tempérance (attitude au pardon, modestie, prudence, autocontrôle) et transcendance (appréciation de la beauté et de l'excellence, gratitude, espoir, humour, spiritualité)²²⁴.

D'après les auteurs, pour chacun d'entre nous, l'exercice de quatre ou cinq de ces vertus clés est associé à un épanouissement personnel et à un sentiment de satisfaction très importants. Chacun aurait un profil de vertus-clé à développer de façon préférentielle. Par exemple, quelqu'un qui est très curieux, aime apprendre, être aimé, et guider un groupe devrait bâtir sa vie de façon à exercer le plus possible ces vertus (plutôt que d'autres). D'autres découvertes intéressantes sont mises en avant : par exemple, il y a des vertus qui se retrouvent plus fréquemment que d'autres parmi les cinq vertus les plus importantes (curiosité, disposition à aimer et être aimé, justice, gentillesse) et d'autres qui contribuent davantage au bien-être personnel, même si elles ne figurent pas parmi les *top five* chez un individu (ex : gratitude). Des interventions de « thérapie positive » visant à développer ces vertus ont montré leur efficacité dans la prévention et le traitement de la dépression et dans la vitesse de guérison d'autres maladies²²⁵.

La classification VIA est un exemple parfait des avantages et désavantages du modèle des traits généraux pour l'étude de la vertu. L'avantage relève de la facilité à produire des résultats intéressants par le simple usage de questionnaires. Certes, la validation et la diffusion de questionnaires qui cherchent à mesurer le niveau des vertus demandent une quantité énorme de travail. Mais c'est une procédure très bien comprise qui relève des compétences classiques de la psychologie de la personnalité. Une fois que les questionnaires existent (et c'est le cas pour les 24

²²⁴ Wisdom and knowledge (creativity, curiosity, judgment/ critical thinking, love of learning, perspective), courage (bravery, industry/perseverance, authenticity, zest), humanity (kindness, love/intimacy, social intelligence), justice (citizenship/teamwork, fairness, leadership), temperance (forgiveness/mercy, modesty/humility, prudence, self-control) and transcendence (awe/appreciation of beauty and excellence, gratitude, hope, playfulness, spirituality).

²²⁵ C. Peterson, N. Park, and M. E. P. Seligman, "Greater Strengths of Character and Recovery from Illness," *The Journal of Positive Psychology* 1, no. 1 (2006), S. Seligman, "Park, & Peterson (2005) Mep Seligman, Ta Steen, N. Park and C. Peterson, Positive Psychology Progress: Empirical Validation of Interventions," *American Psychologist* 60 (2005).

vertus du modèle VIA), il devient très facile d'explorer les corrélats de différents niveaux de vertu et l'efficacité des interventions visant à développer le caractère. Après des millénaires de discussion philosophique, il est sans doute rafraîchissant d'avoir à disposition un flux de données empiriques fiables sur les effets des vertus et les moyens de les développer.

Cependant, l'enthousiasme tout à fait justifiable ne doit pas cacher les limites du projet VIA. Du point de vue d'une compréhension profonde du fonctionnement de la vertu, la classification VIA est très peu ambitieuse. La réflexion philosophique sur la vertu a envisagé l'existence de traits de caractères fiables à l'échelle individuelle et qui interagissent d'une façon très riche entre eux, pour orienter conjointement le comportement et les réactions émotionnelles de l'agent. Non seulement l'attitude d'un individu courageux devrait être cohérente et fiable, mais le courage interagit avec la justice ou la sagesse d'une manière précise et informative.

Or, dans le modèle contemporain des traits on retrouve des tendances peu fiables, étudiées à l'échelle de la population (et non de l'individu) et dont l'interaction est encore moins fiable que l'effet de chaque trait pris séparément. Par exemple, deux personnes très curieuses d'après le modèle VIA peuvent afficher des thèmes situationnels de curiosité très différents : l'un peut être curieux surtout au travail, uniquement pour des problèmes précis et bien définis, et pas dans la vie privée ; l'autre peut être curieux par rapport à sa vie privée, pendant des interactions sociales, et s'intéresser aux réponses possibles aux grandes questions portant sur le sens de la vie. Dans le modèle VIA, cette richesse de détails est perdue dans le score qui résume le niveau de curiosité. Les corrélats de la curiosité sont étudiés à l'échelle de la population : la moyenne efface les différences possibles au niveau de sous-types. Peut-être, deux formes différentes de curiosité n'auront pas les mêmes effets sur le bonheur et la santé. En ce qui concerne l'interaction entre les vertus VIA, la non fiabilité de chaque vertu est multipliée. L'interaction moyenne entre deux effets moyens cache encore plus de sous-types et de détails psychologiques que les phénomènes de départ. Il est possible qu'une certaine forme de curiosité interagisse avec une certaine attitude à être aimé, pour donner des résultats spécifiques : tout cela est perdu dans le modèle VIA.

Jusque-là, il s'agit de problèmes hérités directement de l'approche des traits. Mais l'abandon du modèle lexical et de l'analyse factorielle en faveur de la recherche d'une liste universelle de traits vertueux empire la situation. En fait, malgré toutes les critiques possibles, les chercheurs dans la tradition du *big five* se donnent les moyens pour explorer la *structure* psychologique des traits. L'analyse factorielle montre que certains traits co-occurrent et ont tendance à se présenter ensemble. Les traits de premier ordre sont associés dans le cadre de dimensions ou traits de haut niveau pour cette raison. Par contre, dans le modèle VIA les vertus sont associées en super-traits sur la base de leur objet ; par exemple, toutes les vertus qui concernent la connaissance sont associées

dans la vertu de haut niveau Sagesse et Connaissance. Cela ne prouve pas que la curiosité, l'envie d'apprendre et le jugement critique ont tendance à se développer ensemble. En fait, toute la recherche VIA se développe directement au niveau des 24 vertus et les super-traits ne jouent pratiquement aucun rôle. Dans la liste de vertus universelles, l'équipe de Peterson et Seligman s'est limitée à élaguer les doublons et les candidats trop spécifiques à une culture donnée. Les ambitions théoriques de la théorie de la vertu, ainsi que celles de la psychologie de la personnalité (mettre au jour la structure du caractère), sont complètement abandonnées. Ci-dessus, nous avons opéré une distinction entre d'un côté une littérature psychologique qui étudie les vertus prises séparément, et en dehors d'un modèle général du fonctionnement du caractère, et, de l'autre, la tentative de modéliser le caractère vertueux. La classification VIA se présente comme un modèle de la structure des vertus, mais, au bout du compte, il n'y a que la façon de mesurer les vertus par des questionnaires qui est standardisée : pour le reste, l'approche est aussi vague sur la structure générale de la vertu que la littérature du premier type.

Malgré tout cela, les résultats de la recherche VIA restent très intéressants. Par exemple, les effets de la gratitude sur la satisfaction personnelle et le bonheur n'avaient été jamais prouvés de façon systématique. Même s'il s'agit d'une notion générique et peu fiable de « gratitude », cela reste un résultat important qui peut amener à des interventions efficaces : des thérapies positives visant à augmenter les manifestations de gratitude peuvent générer une amélioration de la satisfaction personnelle à l'échelle d'une population. Ne serait-ce qu'il y a quatre ou cinq ans, des données empiriques de cette qualité sur la vertu n'existaient pas. Toutefois, le succès de la classification VIA ne justifie pas l'abandon systématique des ambitions de la tradition philosophique et psychologique, non sans avoir exploré des modèles alternatifs. Les prochains chapitres seront consacrés à l'exploration d'approches psychologique de la modélisation de la personnalité et de la vertu qui sont minoritaires, mais plus prometteurs.

Approches typologiques et développementales à l'étude de la personnalité

Introduction et résumé

Dans ce chapitre, nous commençons à explorer les approches minoritaires en psychologie de la personnalité. Nous cherchons un interlocuteur qui respecte les *desiderata* de la théorie de la vertu : l'étude, à l'échelle individuelle, de la structure psychologique, de l'interaction et des effets de traits fiables, définis de façon informative et rigoureuse. Dans ce cadre, nous chercherons à réhabiliter le projet d'une approche typologique raffinée de la personnalité et à souligner l'importance d'une approche développementale, qui suit l'évolution de la personnalité au fil de la vie.

Avant tout, nous visons la définition d'un cadre théorique idéal : la formulation d'une typologie « molaire » et développementale de la personnalité. Ce label désigne un genre de psychologie de la personnalité qui en est à ses balbutiements aujourd'hui, mais dont le développement est une condition indispensable au dialogue entre psychologie et théorie de la vertu. Après la définition du cadre conceptuel, le prochain chapitre introduira les théories qui s'approchent le plus de ces exigences méthodologiques.

La notion de type et la stratégie consistant à classer les personnes en plusieurs types ont une histoire aussi vieille et importante que celle de la notion de trait. Malgré cela, la tradition typologique a été nettement tenue à l'écart de la synthèse scientifique de fin 19^{ème} début 20^{ème} qui a marqué la naissance de la psychologie de la personnalité. On peut citer deux raisons. D'abord, aussi bien dans le sens commun que dans les approches théoriques, il semble que la notion de type ou de stéréotype n'ajoute rien au contenu descriptif et explicatif d'une liste de traits. Un type n'était-il pas défini par l'ensemble des traits qui le caractérisent ? Un scientifique « typique » est génial, un peu empoté, déterminé mais un peu distrait, etc. Il semble que le type ne fasse que résumer une description par les traits, sans rien ajouter. Ensuite, les typologies ont aussi une mauvaise réputation, elles ont été longtemps simplificatrices et une source de discrimination.

Ces deux arguments ne sont plus tenables. Comme nous l'avons montré dans le chapitre sur le sens commun, dans l'usage quotidien, les stéréotypes tels que « le scientifique », « le vieux », etc. ont un contenu descriptif plus riche que les traits qui y sont normalement associés. Par exemple, les gens ont une représentation cognitive spécifique de la façon dont une personne âgée est lente, qui n'est pas inférée du trait « lent ».

En ce qui concerne la théorie des traits, récemment les « types » y ont retrouvé leur place. À l'origine, le problème était le suivant : on pourrait postuler l'existence de types uniquement si le niveau des traits dans la population n'était pas distribué de façon normale, mais se concentrait sur les valeurs correspondantes aux types. Par exemple, s'il y a cinq types, les cinq niveaux les plus répandus du trait « extraversion » devraient correspondre respectivement aux cinq niveaux d'extraversion associés aux types, si bien que l'extraversion n'aurait pas une distribution normale dans la population. Normalement, l'approche des traits classique ne prévoit pas « d'effet attracteur » de certains types par rapport à la distribution normale des traits : dans ce cas de figure, la notion de type n'aurait aucun intérêt théorique profond. Un label de type serait juste un nom donné à une distribution possible des traits, déjà prévue par la théorie. Toutefois, l'existence d'effets d'attraction a été récemment démontrée. Cela signifie que les traits s'associent dans un nombre limité de combinaisons beaucoup plus fréquentes que les autres : aujourd'hui, on peut parler d'une typologie basée sur les traits généraux.

Des travaux récents ont mis au jour cette typologie au niveau « macro » (types formés par des traits généraux), mais si on veut (1) dépasser les limites de la théorie des traits – sans tomber dans une démultiplication indéfinie et ingérables des types (typologie micro ou moléculaire) – et (2) éviter les risques d'une typologie trop simplificatrice, il faut chercher à développer une typologie « molaire », c'est-à-dire une typologie définie par l'action de stratégies comportementales moins générales, mais plus fiables. Par exemple, l'idéal serait de découvrir et d'étudier expérimentalement *plusieurs* formes d'extraversion, de timidité ou d'ouverture d'esprit. L'interaction de ces « traits » plus spécifiques définirait des types humains à un niveau molaire.

Pour dépasser vraiment les limites de l'approche des traits généraux, il faut également poser le problème de la continuité et du changement de la personnalité au fil de l'âge, au-delà de la vision limitative du *big five*, selon laquelle la personnalité serait fixée au début de l'âge adulte. En ce sens, élaborer une typologie molaire n'est pas suffisant, si on la conçoit uniquement de façon synchronique : il faut aussi étudier les dynamiques de stabilisation ou d'évolution temporelle de ces types en d'autres types. Quel potentiel d'évolution pour la timidité de quelqu'un qui a un niveau élevé de névrotisme ? Va-t-elle évoluer dans une autre forme de timidité, ou bien s'estomper ? C'est le genre de questions auxquelles une science de la personnalité devra s'attaquer.

Le volume de travail déjà produit dans cette approche est extrêmement limité : des courants minoritaires au sein de la psychologie de la personnalité commencent à peine à devenir sérieusement typologiques et sérieusement développementaux. Pour juger de son succès, il faudra attendre le développement ultérieur de la recherche. Cependant, à l'état actuel les arguments motivant une recherche dans cette direction paraissent contraignants. Le prochain chapitre sera

consacré aux théories qui s'approchent le plus des exigences méthodologiques d'une typologie molaire et développementale.

Approche typologique à la personnalité et théorie de la vertu

L'étude des traits au niveau d'une population est radicalement inadéquate pour classifier et décrire la structure réelle de la personnalité des individus. Quel enseignement faut-il en tirer ? Où faut-il chercher les bases d'une compréhension théorique adéquate des différences individuelles, qui puisse être à la base d'une théorie normative du caractère ? Si la théorie contemporaine des traits – malgré la proximité entre les ambitions des pères fondateurs du domaine (tels que Allport) et celles des philosophes – ne peut pas offrir l'assistance théorique nécessaire, on peut trouver de l'aide inespérée auprès d'une autre approche théorique très classique qui, tout en ayant survécu, a été exclue de la synthèse élaborée par la psychologie de la personnalité du vingtième siècle : l'approche typologique.

La référence à l'existence et à l'action des traits de caractère n'est pas le seul élément central dans les théories spontanées du sens commun. Une autre pratique très répandue est la classification des personnes en « types ». Le nombre immense d'étiquettes sociales mobilisées dans la vie quotidienne témoigne de l'étendue du phénomène : nous disons de quelqu'un qu'il est « un hyper-actif », « un émotif », etc. Il est clair que la liste des « types » mobilisables par cette pratique d'attribution n'est pas moins longue que la liste des traits, car tout terme de trait a un type « équivalent » : « extroverti » / « un extroverti », « hyper-actif » / « un hyperactif », « timide » / « un timide ». En même temps, le contenu exprimé par un terme de type est loin d'être équivalent à « quelqu'un qui possède le trait correspondant » (« un hyperactif » = quelqu'un qui est hyper-actif). Des expériences ont montré que chacune des étiquettes sociales que nous utilisons représente le placement du sujet dans plusieurs dimensions et constitue déjà un petit profil²²⁶. Il s'agit de « *prototypes dérivés de plusieurs caractéristiques formant un schéma cohérent de caractéristiques habituelles*²²⁷ ». En d'autres termes, comme nous l'avons déjà expliqué dans le chapitre sur la théorie du caractère du sens commun, le stéréotype est plus riche en information que le trait correspondant. Le trait de l'« hyper-activité » est probablement une abstraction à partir de représentations très riches en détail

²²⁶ N. Cantor et W. Mischel, "Prototypes in person perception," *Advances in experimental social psychology* 12 (1979), C. N. Macrae et G. V. Bodenhausen, "Social Cognition: Thinking Categorically about Others," *Annual Review of Psychology* 51, no. 1 (2000).

²²⁷ Julien Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" (Thèse de doctorat, 2003) 37.

(façon de parler, de bouger, etc.) de plusieurs hyperactifs typiques. En plus, il y a un grand nombre de types qui ne correspondent pas à un trait particulier, mais qui sont associés à des phases de la vie (le jeune naïf, l'adolescent rebelle, le vieux sage, etc.) et à des rôles sociaux (le nouveau riche, l'avocat arriviste) ou relationnels (le mari infidèle, le père ambitieux, etc.)

Tout comme celle de trait, la notion de type humain n'a pas manqué de stimuler les théoriciens qui ont essayé d'élaborer une compréhension systématique de la personnalité humaine. L'histoire des typologies de personnalité et des types humains est bien longue : Théophraste, disciple d'Aristote, avait repéré 30 types de caractère typiques et exhaustifs dans la société de l'époque, suivant la façon de se conduire en société et de gérer les affaires²²⁸. Plus tard, Hippocrate a proposé une classification en quatre types de caractère, à partir d'une théorie physiologique qui voit l'action de quatre fluides corporels (la bile jaune, la bile noire, le sang et le phlegme) et des quatre éléments de la nature (le feu, l'air, la terre et l'eau) réguler l'équilibre psychique et expliquer les différences individuelles dans des attitudes comme la rationalité, la moralité et les passions. Développant cette théorie, Galen a proposé neuf types tempéramentaux : dans le type idéal, tous les éléments étaient en équilibre, puis il y avait quatre types négatifs dans chacun desquels un seul couple formé par un fluide corporel et un élément naturel avait le dessus. Enfin, dans les quatre types restant – les plus célèbres – deux fluides corporels exerçaient l'influence principale : il s'agit des mélancoliques (qui sont maussades, tristes et déprimés), les sanguins (qui sont très actifs et enjoués), les colériques (qui sont susceptibles, irritables et agressifs) et les flegmatiques (qui sont apathiques et paresseux)²²⁹. Cette histoire est loin de s'arrêter avec les Grecs : des typologies de personnalité – célèbres ou moins célèbres, respectées ou qui suscitent aujourd'hui un peu d'ironie – ont jalonné l'histoire de la pensée occidentale. Même des figures centrales telles que Darwin ne se sont pas soustraites à l'exercice de proposer une typologie humaine²³⁰. En ce qui concerne le vingtième siècle, on peut citer quelques typologies qui ont retenu davantage l'attention des chercheurs : la célèbre théorie de Sheldon (de la moitié du siècle) qui considère que la personnalité est liée à des types corporels (somatotypes)

²²⁸ J. Rusten, I. C. Cunningham et A. D. Knox, "Theophrastus: Characters, mimes, cercidas, and the choliambic poets," (Cambridge, MA: Harvard University Press, 1993).

²²⁹ J. Kagan, *In: Galen's prophecy: temperament in human nature* (New York: BasicBooks, 1994).

²³⁰ R. Wright, *The moral animal: evolutionary psychology and everyday life* (Pantheon Books, 1994).

« Les ectomorphes (grands et minces) seraient réservés, inhibés et craintifs. Les endomorphes (petit et gras) seraient familiers, affectueux et faciles d'approche. Enfin, les mésomorphes (ossature et musculature fortes) seraient quant à eux énergiques, très actifs et confiants²³¹. »

La typologie d'Adler :

« Les dominants manifesteraient un niveau moyen d'intérêt social et un haut niveau d'activité, qui s'exprimerait souvent d'une façon asociale et agressive; les cas extrêmes de ce type deviendraient des criminels. Les accapareurs manifesteraient peu d'intérêt social et un niveau moyen d'activité; ces gens seraient souvent perçus comme étant dépendants des autres. Les évitants afficheraient beaucoup d'intérêt social combiné à un niveau très bas d'activité. Finalement, les bons gars manifesteraient beaucoup d'intérêt social combiné à un niveau d'activité très élevé; ils seraient des individus qui essaient de vivre leur vie pleinement, tout en étant concernés par le bien-être d'autrui. »

Il existe plusieurs typologies d'orientation psychanalytique, dont celles de Freud, Jung, Horney, Reich et Kernberg. Celle en 16 types de Myers-Briggs, basée sur les travaux de Jung, est plutôt populaire²³². La distinction entre les individus de type A et de type B développée par Friedman et Rosenman est aussi célèbre :

« Les individus de Type A sont compétitifs, travailleurs, impatientes, impulsifs, agressifs verbalement, manifestent une forte propension à l'hostilité et aiment prendre des risques. Les individus de Type B présentent le profil contraire, c'est-à-dire qu'ils sont peu compétitifs, patients, peu hostiles ou colériques²³³. »

Malgré cette prolifération, la tradition typologique a été nettement tenue à l'écart de la synthèse scientifique de fin 19^{ème} début 20^{ème} qui a marqué la naissance de la psychologie de la personnalité. On peut dire que, après un tête-à-tête millénaire entre explication par les traits et explication typologique dans le sens commun, la littérature et plusieurs traditions théoriques, l'explication typologique a perdu à la dernière minute, lorsqu'il s'agissait de donner naissance à une véritable science de la personnalité. Pourquoi a-t-il été ainsi ? Une explication superficielle pourrait souligner

²³¹ Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" 39, W. H. Sheldon, "The varieties of human temperament" (New York: Harper Row, 1942).

²³² D. Keirse et M. M. Bates, *Please Understand Me: Character & Temperament Types* (Prometheus Nemesis Book Co., 1984).

²³³ M. Friedman et R. H. Rosenman, *Type A Behavior and Your Heart* (Knopf, 1974), Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" 39.

la faible valeur scientifique des typologies citées par rapport au niveau des standards actuels de validations et répliques, ou le fait qu'une typologie de ce genre pourrait favoriser la diffusion de stéréotypes réducteurs. S'il peut y avoir un peu de vérité à cela, accepter cette explication serait voir les choses à l'envers : en ce qui concerne la période antérieure au début du siècle, le manque de rigueur était un point commun à la tradition typologique et à celle des traits. Pour la période successive, c'est plutôt le manque d'intérêt et le très faible volume de recherche sérieuse dans la tradition typologique qui expliquent le caractère non convaincant et non suffisamment répliqué des résultats, non pas le contraire.

En réalité, il existe une explication théorique profonde de cet abandon de l'approche des types, liée notamment à la question de l'homogénéité locale. Nous avons vu que la psychologie de la personnalité a fortement voulu (et cru pouvoir) étudier des dynamiques psychologiques actives à l'échelle de l'individu et, étant donné qu'elle a procédé à une étude de statistiques des groupes et populations, elle a cru étudier des processus localement homogènes. Dans ce contexte, la notion de type perd toute spécificité théorique et peut être complètement réduite à l'action de plusieurs traits. Si on part du constat que l'appartenance à de différentes catégories (quelles qu'elles soient) d'une typologie doit être justifiée par l'action de facteurs psychologiques de plus bas niveau par rapport au type (en d'autres termes, celle de type ne peut pas être une notion primitive de la théorie, il faut bien définir un type à partir de l'action de processus plus basiques), alors de deux choses l'une : soit l'action de ces facteurs plus basiques est complexe, difficile à résumer et non équivalente dans des individus différents, et alors il est bien utile et informatif de résumer dans des « types » l'effet émergent de ces facteurs. Soit la contribution et l'action de chaque facteur est bien comprise et est la même dans chaque individu, mais alors il vaut mieux s'exprimer directement en mobilisant ces facteurs pour expliquer l'effet global : la catégorisation en type devient parfaitement épiphénoménale.

Or, étant donné que dans l'explication traditionnelle ce facteur plus basique qui explique l'émergence du type est bien sûr le trait, il est clair qu'une théorie scientifique des traits n'a pas besoin de résumer ses résultats par l'élaboration d'une typologie. Par exemple, dans le cas du *big five*, si on peut dire qu'un individu a un haut niveau d'extraversion, névrotisme et un bas niveau de contrainte, pourquoi donner un nom à cette configuration ? Et même si on procédait ainsi par commodité, quelle serait la profondeur théorique et l'utilité d'une telle classification ? Certes, on peut toujours mobiliser des types pour signaler des effets intéressants, par exemple appeler certains individus « équilibrés », mais tout le travail explicatif sera fait par la configuration de traits sous-jacente.

Le fait que le fonctionnement d'un type soit entièrement réductible à l'action de plusieurs traits ne constitue pas la seule raison pour mettre au deuxième plan la notion de type. Dans la théorie traditionnelle, les traits (le positionnement individuel dans la dimension de chaque trait) sont censés être distribués de façon normale dans la population. On ne prévoit pas « d'effet attracteur » de certains types par rapport à la distribution normale des traits. En fait, s'il y a des types de personnalité mutuellement exclusives et exhaustives, disons A, B et C, dans lesquels chaque trait donne une contribution spécifique (l'extraversion est à un niveau très bas x dans le type A, très haut y dans le type B et haut z dans le type C), on devrait observer une concentration des valeurs du trait sur les trois niveaux correspondants aux types, ce qui irait altérer la distribution normale. Dans ce cas de figure, même si l'appartenance à un type s'explique entièrement par l'action des traits, le fait que la distribution des traits se concentre sur les valeurs correspondant aux types donnerait à ces derniers un rôle théorique intéressant²³⁴. Si l'existence de ces effets est aujourd'hui prouvée, celui de la distribution normale a été un postulat central de la théorie des traits en psychologie.

Par rapport à l'approche standard, la situation change si on abandonne le postulat d'homogénéité locale. Si la personnalité est constituée par des traits qui se comportent de façon différente dans des individus différents, si bien qu'il n'existe pas quelque chose comme *la* contribution typique de la timidité à l'économie de la personnalité, alors le type redevient l'unité de mesure centrale de la personnalité.

²³⁴ Cet effet d'attraction des types sur la distribution des traits – qui restent conçus comme universels et centraux dans la théorie – est très intéressant par rapport à la théorie philosophique de la vertu, car il semble la caractériser fort bien. D'un point de vue conceptuel, on a vu qu'une priorité excessive accordée aux traits rend inutile la notion de type. Dans la tradition de l'éthique de la vertu, on insiste beaucoup – et en même temps – sur l'action des différentes vertus et sur des types tels que l'individu vertueux, incontinent ou bestial (par exemple, chez Aristote). Même si les philosophes ne sont pas explicites sur des questions de distribution statistique – et donc on n'est pas sûr du niveau de représentation des cas intermédiaires (entre les types) dans la population (d'après eux) – ils pensent que l'action des traits de caractère se concentre autour de certains types. Il y a donc là une différence claire entre la théorie psychologique et philosophique traditionnelles : la théorie de la vertu est déjà nettement une typologie, même si – comme on le verra – du mauvais genre, car elle postule des types universels et au niveau « macro » de généralité.

Prenons l'exemple d'un groupe de soldats qui se sont distingués pour une conduite exemplaire et des actes de vaillance, et supposons que ce genre de conduite remarquable soit typique et caractérise profondément leur personnalité. Dans l'approche méthodologique de la théorie classique des traits, on essaierait d'expliquer ce tempérament à partir de l'interaction de plusieurs traits généraux : disons le courage, la justice et la détermination. La théorie n'affirme pas que la contribution de chaque trait est toujours la même, de façon mécanique. Au contraire, comme nous l'avons dit, l'existence d'interactions est un aspect central : le courage ne s'exprime pas de la même façon suivant que la personne soit juste et déterminée ou non²³⁵. Mais ces types d'interactions sont censés être universels (le courage et la justice interagissent de la même façon dans tous les individus qui sont à la fois courageux et justes) et prévues par la théorie. En ce sens, si on crée l'étiquette « combattant vaillant » on n'a fait que trouver un nom pour une configuration donnée de traits (l'interaction entre courage, justice et détermination). On est libre de le faire, mais on n'aura pas contribué de manière intéressante à l'explication du comportement des sujets. Dans une théorie qui abandonne l'homogénéité locale et prévoit des phénomènes interactifs beaucoup plus complexes (non pas *le* courage, *la* justice et *la* détermination, mais des formes spécifiques de ces traits interagissant d'une façon unique), postuler un type devient essentiel. Par « soldat vaillant », on pourrait désigner un « écosystème personnel » unique, au sein duquel les différents éléments n'agissent pas d'une façon facilement prévisible par la théorie générale.

La nécessité d'une typologie « molaire »

Pour résoudre le problème de l'approche traditionnelle en psychologie des traits – qui n'est pas celui de postuler des entités inexistantes (Molenaar, Borsboom) mais plutôt celui de ne pas avoir les moyens d'étudier le comportement des traits *chez l'individu réel* – il semble qu'on ait besoin de ce qu'on pourrait appeler une « typologie molaire de la personnalité ».

Pour rester dans le cadre d'une classification scientifique respectable, s'il faut éviter l'extrême de traits généraux et universels associés toujours à la même contribution relative (ce qui permettrait d'élaborer une typologie au niveau macro), il faut éviter aussi un particularisme radical quant au fonctionnement de la personnalité. Dans le jargon technique de la psychologie de la personnalité, la deuxième option est appelée « approche idéographique radicale » : ici on pourrait la définir

²³⁵ Il faut reconnaître que la théorie psychologique des traits *mainstream* est au bout des comptes incapable d'étudier *vraiment* ces interactions : non pas parce qu'elle ne les postulerait pas, mais parce qu'elle les étudie uniquement chez l'*individu moyen*.

« typologie moléculaire (ou micro) de la personnalité ». Cela correspond à un scénario dans lequel les niveaux de spécificité et de particularisme dans le fonctionnement des traits sont si poussés qu'on retrouve des structures totalement uniques et spécifiques à l'individu étudié. La façon tout à fait unique d'être courageux de Pierre interagit avec sa façon aussi unique d'être juste et déterminé. Ensemble, elles expliquent sa conduite. Ce genre d'interaction pourrait à son tour être unique au cas de Pierre : chez quelqu'un qui aurait des traits très semblables à ceux de Pierre, l'interaction pourrait ne pas se produire. Or, avancer une telle position serait équivalente à admettre qu'en réalité il n'y a pas de traits et qu'aucune généralisation scientifique intéressante sur la structure des différences individuelles n'est possible. Rien pour l'instant ne nous contraint à cette conclusion radicale et pessimiste.

L'espoir est celui de pouvoir découvrir une « typologie molaire » dans le cadre de laquelle un degré limité de variation dans le fonctionnement des traits (plusieurs variants informatifs et « isolables » d'un trait donné, sans avoir à attribuer « à chacun sa forme de courage ») donne naissance à une liste informative de types humains. Certes, le besoin de systématisation propre à la science nous pousse à chercher autant de généralités que possible. L'objectif minimal est d'éviter le particularisme radical en élaborant une liste de types validés par les meilleurs standards psychométriques, mais il serait bien sûr très souhaitable de découvrir des liens, des relations systématiques et hiérarchiques entre ces types.

La qualité principale de la théorie traditionnelle des traits réside dans son approche hiérarchique qui assigne une place claire à tout type de trait : dans le *big five*, chaque individu a un niveau défini dans chacune des cinq dimensions générales. Occuper une place donnée dans la dimension de (disons) l'extraversion signifie avoir l'une ou l'autre des distributions possibles au niveau des différentes facettes (sociabilité, confiance en soi, expressivité, énergie, activité et recherche d'excitation). Ensuite, le niveau des facettes est associé à certaines réponses habituelles, dans une structure ordonnée où on sait à quel niveau placer tout phénomène qui entre dans l'économie du caractère. Or, ce serait dommage de devoir reconnaître qu'il existe uniquement un bon nombre de types à un niveau molaire de généralité, sans que ces types entretiennent des relations intéressantes entre eux. Une typologie molaire ne sera certainement pas si structurée et hiérarchique que le *big five* ou les typologies macro de l'antiquité, où les types étaient mutuellement exclusifs²³⁶ et établis sur la base de principes clairs de classification. La recherche n'est qu'à ses débuts, mais on peut espérer que les types molaires qu'on découvrira (ex : narcissique type a, hyperactif type b, etc.)

²³⁶ Du point de vue d'une science de la personnalité, les typologies mutuellement exclusives ont plus de valeurs que les typologies non exclusives et liées à des contextes déterminés (typologies de clients, d'étudiants, d'artistes)

entreteniront des relations non superficielles entre eux et seront associés à des effets profonds (voir chapitre suivant).

L'idée d'une catégorisation des types humains est associée à des mauvaises connotations : son histoire a été caractérisée par des dérapages et évoque des mauvais souvenirs, étant donné la proximité de la notion de « type humain » avec les pires stéréotypes qui ont circulé dans la culture occidentale. Cependant, il faut avoir le courage et la rigueur intellectuelle pour défendre ce projet contre des rapprochements faciles. L'intention n'est pas celle d'enfermer quelqu'un dans la cage conceptuelle représentée par un ou plusieurs « types humains ». Du côté de la psychologie, en particulier les approches récentes qui ont abandonné le modèle monolithique des traits « larges », universels et localement homogènes, l'arrière-plan idéologique est plutôt la volonté de permettre à l'individu un parcours évolutif le plus ouvert possible, à chaque phase de son développement. Une étude la plus objective possible des structures de personnalité qui caractérisent la condition d'un individu à un moment donné – aussi bien qu'une compréhension des pistes d'évolution les plus probables – ne peut qu'aider, beaucoup plus que rester dans le vague d'une compréhension pré-scientifique du domaine de la variation individuelle.

En plus, il ne faut pas s'imaginer que ces types correspondent à des étiquettes sociales faciles : il s'agirait d'une classification technique de dynamiques psychologiques qui structurent la personne, ce qui ne s'exprime pas en termes de rôles sociaux ou clichés tels que « une mère au foyer typique », que ces clichés soient généraux ou spécifiques (« une mère au foyer divorcée et déprimée typique »). Le fait que la typologie cherchée se situe au niveau « molaire » est en mesure de déjouer le risque d'une fixation de stéréotypes rigides.

Il faut reconnaître que toute science cherche une classification des objets dans son domaine d'étude, pour mieux décrire, expliquer et prédire leur comportement. Ce besoin est largement accepté dans le cas de la psychologie clinique, où des taxinomies telles que la classification DSM sont pratique courante. Mais la résistance dans le domaine du fonctionnement « normal » est forte, ce qui constitue une preuve ultérieure des biais en faveur de l'étude de la dimension pathologique et clinique en psychologie. Ainsi, *« Plutôt que de développer un système de classification des personnes, les psychologues ont développé des systèmes de classification des variables comme les traits de personnalité, les motivations ou les valeurs. Si la biologie était dans la même situation que*

*la psychologie, elle classifierait les caractéristiques des animaux telles que les formes de becs, les plumes, etc., et non pas les animaux*²³⁷. »

Typologies « macro » : les adaptés, le sous-contrôlés et les sur-contrôlés

Heureusement, même si la psychologie de la personnalité commence à peine à devenir sérieusement typologique, il y a eu des développements récents intéressants²³⁸. Pour l'instant, la question de sous-types très spécifiques – dont nous avons anticipé l'importance – commence à peine à être envisagée (voir chapitre sur approches socio-cognitives). On reste souvent à un niveau de classification « macro », ce qui est déjà une façon de s'éloigner de la distribution normale des traits postulée par l'approche classique, car on reconnaît que les niveaux des traits se concentrent autour d'un certain nombre de types discrets. Dans cette catégorisation macro, au moins trois types fondamentaux semblent émerger, à partir d'études différentes : les Adaptés (ou Résilients), les Sous-contrôlés et les Sur-contrôlés.²³⁹

« Les Adaptés, généralement le plus grand groupe dans toutes les études, avec près de 70 % des individus, sont caractérisés par des niveaux élevés de Complaisance, de Contrainte et d'Ouverture à l'expérience, un bas niveau de Névrotisme et un niveau moyen d'Extraversion. Ces individus affichent une adaptation personnelle et sociale moyenne ou supérieure. Les Sous-contrôlés, typiquement 20 % des échantillons étudiés, sont caractérisés par de bas niveaux de Complaisance, de Contrainte et d'Ouverture, un haut niveau de Névrotisme et un niveau moyen d'Extraversion. Ils manifestent un profil de problème d'adaptation caractérisé

²³⁷ Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" 37.

²³⁸ Sans doute aidés par la disponibilité de moyens statistiques beaucoup plus puissants (l'étude du positionnement des groupes et des individus dans l'espace multidimensionnel d'une typologie étant statistiquement très complexe) et par la fin du débat personne-situation, qui a débloqué des énergies considérables pour la recherche. Si on n'est même pas d'accord sur ce qui constitue un trait, comment peut-on étudier des types ? Ce serait comme si on cherchait à classer les animaux sans savoir reconnaître un bec ou des plumes.

²³⁹ J. B. Asendorpf et al., "Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults," *European Journal of Personality* 15, no. 3 (2001), A. Caspi, "Personality development across the life course," *Handbook of child psychology* 3 (1998), R. W. Robins, O. P. John et A. Caspi, "The typological approach to studying personality," *Methods and models for studying the individual* (1998), C. F. M. van Lieshout, G. J. T. Haselager et M. A. G. van Aken, "Generalizability of personality types and subtypes across age, gender, and methods of assessment," *Biennial Meeting of the International Society for the Study of Behavioral Development, Berne, Switzerland* (1998).

par une forte désinhibition dans l'adaptation personnelle et sociale (i.e., impulsivité, agressivité, délinquance, etc.). Enfin, les Sur-contrôlés, généralement le plus petit groupe avec près de 10 % des individus, sont caractérisés par des niveaux élevés de Névrotisme et de Complaisance, des niveaux moyens de Contrainte et d'Ouverture et un bas niveau d'Extraversion. Ces individus manifestent plutôt un profil de problème d'adaptation caractérisé par une forte inhibition dans l'adaptation personnelle et sociale (i.e., retrait social, dépression, anxiété, etc.)²⁴⁰. »

Ces types ont été retrouvés dans plusieurs échantillons, indépendamment de culture, sexe, rôle social et âge. Quoique moins consensuelle, une subdivision en sous-types a déjà été proposée par certains chercheurs :

« Les Adaptés constituent deux groupes distincts, soit les Communales et les Agentifs. Les Communales sont caractérisés par de plus bas niveaux d'Extraversion et de Névrotisme et des niveaux plus élevés de Complaisance et de Contrainte, alors que les Agentifs sont caractérisés par de plus hauts niveaux d'Extraversion, d'Ouverture et de Névrotisme et de plus bas niveaux de Complaisance et de Contrainte. En général, les Agentifs tendent à présenter davantage de problèmes d'adaptation que les Communales, particulièrement des comportements antisociaux. Les Sous-contrôlés constitueraient aussi deux groupes distincts, les Impulsifs et les Antisociaux. Les Impulsifs sont caractérisés par de plus hauts niveaux d'Extraversion, de Complaisance et de Névrotisme et de plus bas niveaux de Contrainte et d'Ouverture, alors que les Antisociaux sont caractérisés par de plus bas niveaux de Complaisance, d'Extraversion et de Névrotisme. En général, les Antisociaux tendent à présenter des comportements antisociaux plus variés, fréquents et violents. Finalement, les Sur-contrôlés constituent aussi deux groupes distincts, soit les Vulnérables affectifs et les Orientés vers la réussite. Les Vulnérables affectifs sont caractérisés par de plus hauts niveaux de Névrotisme et d'Ouverture et de plus bas niveaux d'Extraversion, de Complaisance et de Contrainte, alors que les Orientés vers la réussite sont caractérisés par de plus bas niveaux de Névrotisme et de plus hauts niveaux d'Extraversion, de Contrainte et d'Ouverture. En général, les Vulnérables tendent à présenter davantage de problèmes d'adaptation liés à l'humeur telle que la dépression, alors que les Orientés vers la réussite tendent à présenter davantage des problèmes d'anxiété et liés à la compulsion²⁴¹. »

²⁴⁰ Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" 42. Nous n'avons pas modifié l'usage des majuscules dans le texte cité.

²⁴¹ Ibid. 42-43. Nous n'avons pas modifié l'usage des majuscules dans le texte cité.

Même si nous ne retenons cette classification (sans doute provisoire) que comme un exemple de typologie non réductible aux traits, il paraît que les médecins et philosophes de l'antiquité aussi bien que les pionniers d'une approche typologique au vingtième siècle ne s'étaient pas complètement égarés, car certaines correspondances entre les adaptés (sanguins de Galen, endomorphes de Sheldon, bons gars de Adler, Type B de Friedman et Rosenman), les sous-contrôlés (colériques de Galen, mésomorphes de Sheldon, dominants de Adler, Type A de Friedman et Rosenman) et les sur-contrôlés (le tempérament apathique de Thomas et Chess) peuvent être repérées. Mis à part le problème de développer cette typologie à un niveau beaucoup plus particulariste et molaire, avec des sous-types beaucoup plus spécifiques, cette classification tripartite (que personne ne considère complète, il s'agit plutôt d'affirmer qu'il existe *au moins* ces trois types) est parfaitement « synchronique ». Elle ne s'intéresse pas à la question – essentielle, comme nous le verrons dans la section suivante – du développement des types au fil des âges de la vie. On ne sait pas si l'appartenance à un type reste constante au fil du développement individuel.

Approche développementale à la personnalité

Une caractéristique très importante de la théorie de la vertu est l'attention au développement du caractère au fil du temps. Il faut du temps pour développer la vertu – un long processus d'éducation. Une fois en place, la vertu n'est pas là pour nous abandonner, mais pour guider et façonner notre vie dans sa totalité²⁴². Le cadre théorique sous-jacent au développement de la psychologie de la personnalité était certainement en syntonie avec cette position, mais il faut reconnaître que la dimension développementale a été longtemps presque absente de la pratique expérimentale de la discipline²⁴³. L'idée selon laquelle la personnalité une fois stabilisée ne changerait guère pour le reste de la vie (ce qui paraissait confirmé par les données dans une première phase) a contribué à un manque d'attention pour le côté développemental.

En ce qui concerne la psychologie, le père fondateur de l'étude scientifique de la personnalité et premier théoricien des traits, Gordon Allport, était extrêmement sensible au thème du changement de la personnalité au fil du temps. Même s'il ne renonçait pas à postuler une convergence vers une stabilisation et une sorte de « personnalité mature », il la plaçait à un âge plutôt avancé, point

²⁴² E. Deschavanne et P. H. Tavoillot, "Philosophie des âges de la vie," *Paris, Grasset, 2007*.

²⁴³ En psychologie, il faut toujours garder à l'esprit la différence entre le cadre théorique et la pratique expérimentale qui cherche à rendre opérationnels certains aspects de ce cadre – et évite, consciemment ou pour une question d'opportunités, de moyens et de ressources, d'en opérationnaliser d'autres !

d'arrivée plus près de la mort qu'il ne l'est de la naissance²⁴⁴. En opposition nette avec les théories freudiennes qui dominaient à l'époque, « *selon lui, la personnalité change parce que les personnes ont des intentions conscientes (i.e., des buts ou objectifs de vie) qui se développent indépendamment des besoins biologiques de base. Allport considère que l'être humain adulte est guidé davantage par ses intentions actuelles que par des causes antérieures. Les ambitions, les plans ou les aspirations de la personne modulent ses conduites habituelles et constituent des ensembles cognitifs diversifiés et individualisés*²⁴⁵. »

Cette forme d'ouverture va être quelque peu modulée chez deux autres théoriciens classiques des traits, Cattell et Eysenck. Même si ces auteurs proposent une théorie aussi articulée que celle de Allport, ils insistent plus que lui sur les bases biologiques et physiologiques des traits. Cela les pousse, sur la question du développement, à considérer la maturation biologique comme une étape essentielle. Après la fin de l'adolescence on entrerait dans un registre de stabilité où le changement est plutôt un aménagement²⁴⁶. L'influence de l'environnement pour Cattell se concentre essentiellement dans les premières années de vie, où le processus de socialisation est le plus intense, et cela semble confirmer l'idée d'une stabilisation vers la fin de l'adolescence.

Ensuite, la nouveauté décisive dans le développement de la théorie des traits a été la disponibilité de plusieurs études longitudinales, qui mesuraient la personnalité des mêmes individus à plusieurs étapes de leur vie. Ces études n'étaient pas disponibles pendant la période d'activité d'Allport, Eysenck et Cattell, mais, après la crise situationniste, elles ont commencé à attirer l'attention des principaux théoriciens du renouveau de la théorie des traits. Costa et McCrae, non pas les inventeurs mais ceux qui ont mieux développé et popularisé le modèle en cinq facteurs, avaient

²⁴⁴ La conception d'une « personnalité mature » de Allport est la suivante : « (1) un déploiement de soi étendu (participation à des activités sociales et sociétales, prise d'engagements liés à l'avenir) ; (2) une capacité d'intimité et de compassion pour autrui (capacité de relations chaleureuses avec les membres de la famille et les amis, respect et compréhension des autres) ; (3) une sécurité affective (capacité de garder son calme, de contrôler ses émotions) ; (4) une acceptation de soi (accepte ses limites et ses besoins, ses erreurs et ses frustrations) ; (5) des perceptions réalistes (de soi, du milieu social, des problèmes existentiels à résoudre ; capacité de mettre entre parenthèses ses intérêts personnels et de s'engager dans son travail) ; (6) une connaissance objective de soi et des autres (perception de soi objective, capacité de rire de soi, vision ou philosophie de la vie humaine) » Morizot, « Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes » 22.

²⁴⁵ Ibid. 21.

²⁴⁶ R. B. Cattell, *Description and Measurement of Personality* (World book company, 1946), R. B. Cattell, *Personality and motivation structure and measurement* (World Book, 1957), H. J. Eysenck, *A Model for Personality* (Springer, 1981), H. J. Eysenck, *The Structure of Human Personality* (Methuen, 1970).

initialement proposé une image plutôt statique de la personnalité, plus ou moins fixée et cristallisée dès les 30 ans²⁴⁷. Les données longitudinales accumulées entretemps les ont contraints à moduler leurs affirmations et à reconnaître certains changements :

«*De 18 à 30 ans, il y a une diminution du Névrotisme, de l'Extraversion et de l'Ouverture à l'expérience et une augmentation de la Complaisance et de la Contrainte. Après l'âge de 30 ans, McCrae et Costa précisent que les mêmes tendances développementales sont observées, mais dans une moindre magnitude*²⁴⁸. »

Ces changements, que les chercheurs pensent avoir isolés dans plusieurs langues et cultures grâce aux méthodes longitudinales, représenteraient un processus universel de maturation : une véritable maturation psychologique normative, canalisée par des facteurs essentiellement biologiques liés au processus de vieillissement.

Il est difficile de ne pas remarquer, sans bien sûr en contester la validité scientifique, que ces résultats se rapprochent du lieu commun de l'adulte et de la personne âgée comme moins agités, plus calmes, avec davantage d'auto-contrôle, mais moins ouverts et extravertis que le jeune. Cette concession aux effets du vieillissement paraît minimale : la théorie des traits est fondamentalement incapable d'articuler la dimension du changement au fil de la vie. D'autres approches à l'étude du développement de la personne (qui ne sont pas des théories des traits) ne sont pas avares en étapes, phases et crises d'identités qui surviendraient à plusieurs reprises. Il suffit de penser à la célèbre théorie des stades de la vie d'Erikson et à d'autres qui comme celle-ci sont d'inspiration psychanalytique (extension des phases de Freud), ou à toute une famille de théories inspirées de la sociologie et qui postulent un processus de redéfinition et de transaction continue avec l'environnement²⁴⁹. Les théories du premier genre postulent une série de phases dont l'articulation

²⁴⁷ R. R. McCrae et P. T. Costa Jr, "A five-factor theory of personality," *Handbook of personality: Theory and research* 2 (1999), R. R. McCrae et P. T. Costa, "Toward a new generation of personality theories: Theoretical contexts for the five-factor model," *The five-factor model of personality: Theoretical perspectives* (1996).

²⁴⁸ R. R. McCrae et al., "Nature over nurture: temperament, personality, and life span development," *Journal of Personality and Social Psychology* 78, no. 1 (2000), Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" 24.

²⁴⁹ Pour les théorie des stades : E. Erikson, *Childhood and Society. revised and Enlarged* (New York: Norton, 1963), R. Kegan, *The evolving self: problem and process in human development* (Harvard University Press, 1982), L. Kohlberg, *The Psychology of Moral Development: The Nature and Validity of Moral Stages* (Harper & Row, 1984), D. Levinson, *The Seasons of a Man's Life* (New York: Alfred A. Knopf Inc, 1978), D. J. Levinson, "A conception of adult development," *American Psychologist* 41, no. 1 (1986), J. Loevinger, "Measurement of Personality: True or False," *Psychological Inquiry* 4, no. 1 (1993), J. Loevinger, "Stages of personality development," *Handbook of personality psychology* (1997), J. Loevinger et A. Blasi, *Ego Development:*

ne se réduit pas aux effets prévisibles du vieillissement biologique. Les théories du deuxième genre montrent comment des situations stimulantes, de nouvelles motivations, une évolution des relations ou le changement des rôles sociaux sont toujours en mesure de modifier de façon durable les inclinations de base de la personnalité. Est-ce que, pour des raisons conceptuelles et théoriques de fond, les traits se situent au mauvais niveau d'analyse pour chercher des traces de cette variation ? Ou bien cela ne fait-il que mettre en évidence un problème dans la théorie des traits ? Nous croyons que la deuxième réponse est la bonne : indépendamment des arguments biologiques, l'étude des structures de la personnalité de l'homme moyen (avec des statistiques sur des groupes) au lieu des différents « hommes réels » dans les sous-groupes (effacés par le passage à la moyenne) a certainement masqué plusieurs dynamiques de changement des traits au fil du temps, tout comme elle l'a fait au niveau synchronique.

Plus on s'intéresse à des structures de personnalité spécifiques et fiables, plutôt qu'à des inclinations générales (telles que les super-traits du *big five*), plus le postulat de la stabilité paraît difficilement défendable. S'il est envisageable que les traits – définis comme tendances générales – restent les mêmes, il y a certainement plus de variations au niveau d'attitudes spécifiques et fiables.

Conceptualiser la stabilité et la continuité de la personnalité dans le temps

En psychologie de la personnalité, ce n'est que récemment qu'on a posé des bases solides pour étudier systématiquement la possibilité d'un changement dans l'organisation des traits. On ne peut

Conceptions and Theories (Jossey-Bass, 1976), B. L. Neugarten, "Personality and the aging process," *Gerontologist* 12, no. 1 (1972), B. L. Neugarten, "Time, age, and the life cycle," *Gerontologist* 136, no. 7 (1979), R. L. Selman, *The growth of interpersonal understanding: developmental and clinical analyses* (Academic Press, 1980), G. Valliant, "Adaptation to life," *Boston: Little, Brown* (1977). Les théories transactionnelles : P. B. Baltes, "On the incomplete architecture of human ontogeny. Selection, optimization, and compensation as foundation of developmental theory," *The American Psychologist* 52, no. 4 (1997), P. B. Baltes, "Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline," *Developmental Psychology* 23, no. 5 (1987), Caspi, "Personality development across the life course," A. Caspi, "Personality in the life course," *Journal of Personality and Social Psychology* 53, no. 6 (1987), A. Caspi et T. E. Moffitt, "When Do Individual Differences Matter? A Paradoxical Theory of Personality Coherence," *Psychological Inquiry* 4, no. 4 (1993), J. A. Clausen et C. J. Jones, "Predicting Personality Stability Across the Life Span: The Role of Competence and Work and Family Commitments," *Journal of Adult Development* 5, no. 2 (1998), J. S. Clausen, "Adolescent Competence and the Shaping of the Life Course," *The American Journal of Sociology* 96, no. 4 (1991), G. H. Elder Jr, "The Life Course as Developmental Theory," *Child Development* 69, no. 1 (1998), N. Haan, R. Millsap et E. Hartka, "As time goes by: change and stability in personality over fifty years," *Psychology and Aging* 1, no. 3 (1986).

définir le changement que sur la base d'une conceptualisation claire de la continuité. Au moins cinq formes non équivalentes de continuité ont été distinguées. Il s'agit de la continuité structurelle, la continuité différentielle, la continuité absolue, la continuité centrée sur la personne ainsi que la cohérence²⁵⁰.

La *continuité structurelle* se réfère au maintien de la même structure de base dans l'organisation de la personnalité : par exemple, dans le cas du *big five*, cela signifie simplement qu'à des âges différents, la personnalité continue d'être analysable en cinq facteurs. La question paraît anodine, car, intuitivement, les chercheurs et le sens commun ont tendance à croire que la personnalité est caractérisée ou non par une certaine structure de base, tout court. Si les différents types de traits qui s'organisent en cinq dimensions fondamentales décrivent de façon satisfaisante une personne à un moment donné, il paraît naturel de penser que la même structure conceptuelle sera adéquate après 20 ans. Mais en fait, il s'agit d'un postulat qui doit être démontré, et, avant les études longitudinales récemment disponibles, bien peu avait été fait pour prouver cette forme essentielle de continuité. Cela est étonnant, car il s'agissait d'un véritable point faible dans la construction théorique de la théorie des traits. S'il n'y a pas de continuité structurelle, on n'a pas le droit d'interpréter les corrélations entre les super-traits du *big five* et d'autres données sur la vie des sujets comme une forme de validation du modèle. On ne peut pas affirmer que les extrovertis et consciencieux ont davantage de succès au travail dans une phase avancée de leur carrière, si on ne sait pas déjà qu'à 40 ans la personnalité est toujours organisée en cinq facteurs (par rapport à un âge antérieur). Pour tester la continuité structurelle il est nécessaire de vérifier, à partir des résultats des tests sur des échantillons représentatifs de la totalité de la population d'un certain âge x , si la matrice de corrélations qui tient ensemble le modèle est à peu près la même. Il faut que les items de base qui correspondent aux réponses habituelles relatives à un trait de premier ordre soient corrélés plus au moins de la même manière à des âges différents, et ainsi de suite pour la corrélation des traits de premier niveau qui déterminent les super-traits d'ordre supérieur.

Le deuxième type de continuité, la *continuité différentielle*, est le plus étudié en psychologie. Il s'agit d'une mesure de la position relative (sur l'échelle d'un trait) d'un individu par rapport à un groupe, indépendamment de la valeur absolue qu'on pourrait attribuer à certaines actions. Par exemple, quelqu'un peut être le plus timide d'un groupe, et rester tel même si tout le groupe (et lui aussi) devient beaucoup plus extraverti avec l'âge sur une échelle de référence absolue. Inversement, quelqu'un peut changer d'un point de vue différentiel même si, dans l'absolu, il garde

²⁵⁰ A. Caspi et B. W. Roberts, "Personality Development across the Life Course: The Argument for Change and Continuity," *Psychological Inquiry* 12, no. 2 (2001).

la même attitude de toujours (le groupe change alors qu'il reste le même). La recherche a été biaisée en faveur de l'étude de cette forme de continuité parce qu'elle est plutôt simple à analyser d'un point de vue statistique. Les résultats montrent une stabilisation progressive des différences relatives entre les individus, tendance qui atteindrait un plateau vers la soixantaine²⁵¹.

Le troisième type de continuité est l'inverse de la précédente : la *continuité absolue*. L'idée est conceptuellement très simple : garder la même personnalité et la même attitude, mesurées sur des échelles absolues. Du point de vue opérationnel, elle est moins immédiate à calculer que la continuité relative, et elle a été aussi un peu moins étudiée. Les études longitudinales récentes qui la prennent en compte confirment l'image d'une maturation progressive jusqu'à la soixantaine avancée. Comme nous aurons moyen de montrer par la suite, la distinction entre continuité relative et absolue est très intéressante pour la théorie de la vertu, car, si la psychologie de la personnalité a été plutôt biaisée en faveur de la première, la théorie de la vertu se concentre résolument sur la deuxième.

Parmi les notions de traits utilisées par le sens commun, il y en a certaines qui correspondent à des traits absolus et d'autres à des traits relatifs. Si on s'intéresse à l'irascibilité, on pense plutôt à une personne qui va réagir de façon plus violente et émotionnellement chargée que la moyenne lorsqu'elle se sent menacée: ce sera donc un trait avant tout comparatif. Par contre, dans une visée éthique, remarquer que tout le monde fait la même chose ne constitue probablement pas une justification recevable pour des actes de violence. Dans cette acception, « violent » est un trait non-comparatif ou absolu. L'intérêt de la théorie de la vertu pour le développement éthique de l'individu justifie l'attention pour les traits non-comparatifs, alors que l'intérêt des psychologues pour l'interaction sociale standard (profondément conditionnée par des standards relatifs à des groupes ou des contextes sociaux) explique l'étude plus approfondie des traits comparatifs. Il sera important de se souvenir de cela lors de l'évaluation de la pertinence des résultats psychologiques pour une théorie normative du caractère.

Au niveau opérationnel, les trois formes de continuité introduites, malgré l'intention et le cadre conceptuel, sont *estimées* à partir de statistiques de groupes : le même type de statistiques qui ne s'intéressent qu'à l'individu moyen et ignorent la possibilité de sous-groupes. Face à ce problème, Caspi et Roberts postulent une quatrième forme de continuité, qui, théoriquement, aurait dû être déjà comprise dans le cadre conceptuel des trois précédentes. Il s'agit de la continuité *centrée sur la personne* :

²⁵¹ B. W. Roberts et W. F. DelVecchio, "The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies," *Psychological Bulletin* 126, no. 1 (2000).

« Elle se réfère à la continuité intra-individuelle d'une configuration de traits de personnalité à de différents âges. Ainsi, alors que la continuité absolue examine le maintien du niveau moyen d'un trait dans un groupe, la continuité centrée sur la personne examine le niveau moyen de différents sous-groupes d'individus manifestant des trajectoires dans les traits qui sont qualitativement et quantitativement distinctes²⁵². »

Cette dernière forme de continuité est la plus difficile à évaluer, car il s'agit d'une forme de *cohérence* dans le changement. L'intuition est la suivante : un enfant agressif et impulsif qui devient un jeune criminel coupable de violence conjugale manifeste une certaine cohérence de développement. D'un point de vue rigoureux, au niveau phénotypique, il s'agit de traits différents et donc on ne peut pas parler de continuité, qui est toujours « homotypique » (relative au maintien de traits phénotypiques similaires à des âges différents). Pour l'instant, nous laisserons de côté cette forme de cohérence, car être sûrs de voir de la continuité dans la différence présuppose la connaissance des mécanismes profonds qui gèrent la personnalité humaine. À l'état actuel du développement d'une science de la personnalité, on commence à peine à sonder ces mécanismes.

Comme nous l'avons dit, la théorie de la vertu s'intéresse essentiellement aux plus fiables parmi les processus de personnalité réellement actifs dans la psychologie de l'individu, et qui le façonnent au fil du temps. Cela ne signifie pas que des tendances générales, de très faibles propensions « probabilistiques » n'ont aucun intérêt éthique : une tendance faible et générale peut être associée à des résultats spécifiques qu'on ne pourrait pas atteindre avec une attitude plus marquée. Par exemple, une attitude timide générale peut provoquer des réactions non équivalentes à celles produites par des formes spécifique de timidité (voir chapitre sur les approches socio-cognitives). En plus, des tendances de ce genre peuvent être à la base d'effets émergents importants au niveau du fonctionnement de la société. Une tendance générale au courage dans un peuple peut expliquer en partie ses succès militaires. Mais, en général, une théorie normative du caractère *individuel* – qui porte sur le type de personne qu'un individu réel pourrait et devrait devenir – va se concentrer sur les effets les plus profonds, durables et fiables dans le façonnement du caractère.

Dans l'état actuel de la recherche scientifique, le niveau d'analyse de la structure de la personnalité le plus compatible avec ces objectifs est celui d'une typologie *molaire-développementale*. En ce sens, élaborer une typologie molaire n'est pas suffisant, si on la conçoit uniquement de façon synchronique : il faut aussi étudier les dynamiques de stabilisation ou d'évolution temporelle de ces types en d'autres types. Quel potentiel d'évolution pour la timidité de quelqu'un qui a un niveau

²⁵² Morizot, "Le développement de la personnalité de l'homme de l'adolescence au milieu de la vie : Approches centrées sur les variables et sur les personnes" 30.

élevé de névrotisme ? Va-t-elle évoluer vers une autre forme de timidité, ou bien s'estomper ? C'est le genre de questions auxquelles une science de la personnalité devra s'attaquer.

Le volume de travail déjà produit dans cette approche est extrêmement limité : la psychologie de la personnalité commence à peine à devenir sérieusement typologique et sérieusement développementale. Pour juger de son succès, il faudra attendre le développement ultérieur de la recherche. Cependant, à l'état actuel les arguments motivant une recherche dans cette direction paraissent convaincants. Comme nous l'avons dit, si on découvrait encore plus de particularisme que la dose compatible avec une typologie molaire-développementale, aucune science de la personnalité ne serait possible. Dans le chapitre sur le situationnisme, nous montrons que les arguments en faveur de cette fragmentation radicale ne sont pas conclusifs. Inversement, la recherche de structures fiables et très générales a été suffisamment frustrée par des décennies de controverses en psychologie de la personnalité. Notre conclusion est la suivante : si des structures objectives et solides de la personnalité – sur lesquelles une théorie qui a les ambitions de la théorie de la vertu pourrait se baser – existent, c'est dans la direction d'une typologie molaire-développementale qu'il faut les chercher.

Macro-typologies développementales

À l'heure actuelle, il n'existe pas d'approche bien développée qui soit en même temps typologique, développementale et molaire. On trouve des tentatives de construction d'une typologie-développementale qui dépassent le caractère synchronique de la typologie tripartie (adaptés, sous-contrôlés et sur-contrôlés), même si le volume de recherches disponibles dans ce cadre conceptuel est malheureusement très réduit.

L'ambition n'est pas simplement celle d'étudier l'évolution de l'appartenance aux mêmes types qu'on a découverts au niveau synchronique, pour vérifier, par exemple, si des adaptés peuvent devenir sur ou sous-contrôlés. Il s'agit de définir des types qui sont associés à des *trajectoires développementales*. En ce sens, la notion même de similitude des types devient diachronique : deux personnes semblables sont deux individus qui se développent suivant la même trajectoire à des étapes différentes de leur vie. Même si la similitude est circonscrite à certains aspects de leur personnalité, il s'agit d'une similitude dans la façon d'évoluer dans le temps.

Le pionnier d'une typologie développementale a été Block²⁵³. Son étude, très sophistiquée d'un point de vue méthodologique et vraiment en avance par rapport à son temps, est restée

²⁵³ J. Block et N. Haan, *Lives through time* (Bancroft Books Berkeley, 1971).

malheureusement isolée et ses résultats n'ont pas été répliqués. Il se basait sur deux échantillons d'hommes et de femmes, à l'adolescence (14-17 ans) et à l'âge adulte (30-37 ans). Parmi les cinq types identifiés, trois correspondent aux types du modèle tripartite et restent stables (résilients, 30%, sous-contrôlés, 16% et sur-contrôlés, 19%), ce qui représente une réplique des résultats du modèle synchronique à l'intérieur d'un cadre conceptuel plus vaste. Les deux derniers types, qui montrent l'utilité de l'approche et sa supériorité conceptuelle, sont : les adaptés-tardifs (17,5%), qui affichent des problèmes d'adaptation à l'adolescence qui seront ensuite dépassés à l'âge adulte et les extrovertis-anomiques (17,5%), qui suivent le parcours inverse, avec une adolescence normale et des problèmes d'adaptation qui surviennent à un âge adulte.

Les études qui ont suivi l'exemple de Block présentent une variété d'approches méthodologiques non facilement comparables, si bien qu'il est difficile d'arriver à un cadre global. Néanmoins, il semble que quatre profils reviennent de façon générale : (1) un large groupe d'individus bien adaptés, qui restent tels ou voient même améliorer leur condition, la plus favorable en termes d'adaptation personnelle et sociale ; (2) un petit groupe de sous-contrôlés, dont la condition tend à s'améliorer avec l'âge ; (3) un plus petit groupe de sur-contrôlés qui restent par contre plutôt stables dans leur trajectoire, avec des problèmes d'anxiété et de dépression ; (4) le quatrième groupe, les anomiques, est également petit et correspond aux extrovertis-anomiques de Block, avec un processus d'adaptation personnelle qui se dégrade au fil de l'âge.

Il s'agit certainement d'une typologie développementale, mais pas encore « molaire », car le niveau d'analyse choisi est sans doute « macro », la population étant regroupée en trois ou quatre types développementaux. Pour un cadre conceptuel typologique molaire le plus proche possible de la conceptualisation du caractère dont la théorie de la vertu aurait besoin, il faudra explorer les approches socio-cognitives à la personnalité, auxquelles sera consacré le prochain chapitre.

Modèles socio-cognitifs de la personnalité

L'insatisfaction avec presque un siècle de théorie des traits et l'essor contemporain des sciences cognitives ont amené plusieurs théoriciens issus de l'approche traditionnelle à explorer les processus cognitifs à la base de la personnalité individuelle. Ce nouveau courant est appelé approche « socio-cognitive » et cherche à établir une véritable psychologie cognitive de la personnalité. L'inspiration est puisée de travaux de pionniers tels que Bandura et Mischel ; parmi les auteurs contemporains plus significatifs on retrouve Shoda, Cervone, Morf, Kuhl, Cloninger, Carver et Scheier ainsi que les travaux récents de Bandura et Mischel²⁵⁴. Même si ces modèles cognitifs n'en sont qu'à leurs balbutiements au niveau de la conceptualisation et encore plus à niveau des généralisations intéressantes, il paraît évident qu'ils constituent l'avenir du domaine et une étape significative dans le développement de l'étude scientifique de la personnalité. Le choix du niveau d'analyse des modèles socio-cognitifs assure davantage de pertinence pour l'individu, de précision et de fiabilité des prédictions.

Mais le potentiel véritable réside plutôt dans la possibilité d'exploiter le grand nombre d'études sur la cognition sociale qui s'accumulent de façon exponentielle depuis quelques décennies. Il doit forcément être possible d'en profiter pour une compréhension de la structure de la personnalité. Toute distinction entre psychologie de la personnalité et psychologie sociale n'est qu'artificielle : la personnalité des individus est un facteur-clé pour comprendre les dynamiques sociales. Le problème qui se pose est celui de gérer au niveau de conceptualisation « molaire » la tour de Babel de connaissances « moléculaires » sur le fonctionnement cognitif de : schémas, stéréotypes, concepts de trait, standards évaluatifs, systèmes d'auto-régulation, le soi-cognitif etc. La liste des découvertes qui ont vocation à intéresser le théoricien de la personnalité est longue et a tendance à se confondre

²⁵⁴ Pour une revue, voir D. Cervone, "PERSONALITY ARCHITECTURE: Within-Person Structures and Processes," *Annual Review of Psychology* 56, no. 1 (2005). Des articles représentatifs des différents auteurs : C. S. Carver et M. F. Scheier, "Control Processes and Self-Organization as Complementary Principles Underlying Behavior," *The Dynamic Perspective in Personality and Social Psychology* 6, no. 4 (2002), D. Cervone, "The Architecture of Personality," *Psychological Review* 111, no. 1 (2004), C. R. Cloninger, "8 Completing the Psychobiological Architecture of Human Personality Development: Temperament, Character, and Coherence," *Understanding Human Development: Dialogues with Lifespan Psychology* (2003), J. Kuhl et S. L. Koole, "Workings of the Will," *Handbook of Experimental Existential Psychology* (2004), Walter Mischel et Yuichi Shoda, "A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure," *Psychological Review* 102, no. 2 (1995).

avec la table des matières d'un manuel de psychologie sociale cognitive. Essayons de résumer les principaux choix théoriques des psychologues qui ont tenté de mettre de l'ordre dans le domaine.

Une première contrainte de clarté est la précision au niveau des entités théoriques postulées. Une ontologie minimale à respecter par toute approche cognitive est celle qui opère la distinction entre systèmes cognitifs, structures cognitives²⁵⁵ et processus cognitifs. Par exemple, lorsque je raisonne (processus) sur mon plan d'action, des croyances, désirs et valeurs (structures) sont vraisemblablement traités par plusieurs systèmes cognitifs (par exemple, des modules d'inférence). L'avantage des approches cognitives réside dans la possibilité d'explorer de façon indépendante les propriétés et de vérifier l'activation de systèmes et structures cognitifs : pour cela, les postulats de la théorie doivent être clairs. Cette contrainte n'était pas remplie par les premières tentatives pour donner une épaisseur cognitive à la psychologie de la personnalité. Depuis les années 70, Mischel avait proposé cinq variables décisives censées être les unités d'analyse fondamentales d'une théorie cognitive de la personnalité : (1) compétences d'élaboration de pensées et actions ; (2) stratégies et structures d'encodage (*encoding*) ; (3) attentes par rapport au monde et à soi-même ; (4) valeurs subjectives ; (5) systèmes d'auto-régulation et de gestion des plans²⁵⁶. Bandura basait sa théorie sur cinq capacités décisives : (1) la capacité à encoder et traiter de l'information symbolique ; (2) la capacité à acquérir des capacités et de la connaissance par l'observation ; (3) la capacité à anticiper des événements futurs ; (4) des capacités d'auto-régulation dans la gestion de ses propres buts et plans ; (5) des capacités auto-réflexives, notamment évaluer sa propre efficacité dans des contextes d'action (*self-efficacy*)²⁵⁷.

Dans les deux cas, l'intention est évidemment celle d'expliquer des profils de personnalité par des interactions informatives entre ces facteurs clés. Mais « l'ontologie » postulée par la théorie est ambiguë : il n'y a pas de principes généraux pour classer ces unités fondamentales. Il est tout à fait possible que leur rôle soit important, mais chacune des deux listes est hétérogène. Les éléments ne se distinguent pas clairement : chez Bandura, la capacité à encoder de l'information symbolique est à la base des autres, et on ne trouvera pas de raisons claires pour opérer une distinction de principe entre l'auto-évaluation (5) et l'auto-régulation (la première pourrait être une phase spécifique de la

²⁵⁵ Terme général pour désigner chacune des « entités mentales » spécifiques qui peuvent être mobilisées par des systèmes cognitifs dans le cadre de processus cognitifs : croyances, désirs, valeurs, buts, etc. Ma croyance que je suis un individu respectable peut être mobilisée dans plusieurs processus d'inférence et par plusieurs modules d'inférence.

²⁵⁶ W. Mischel, "Toward a cognitive social learning reconceptualization of personality," *Psychology Review* 80, no. 4 (1973).

²⁵⁷ A. Bandura, "Self-efficacy: Toward a unifying theory of behavior change," *Psychological Review* 84, no. 2 (1977).

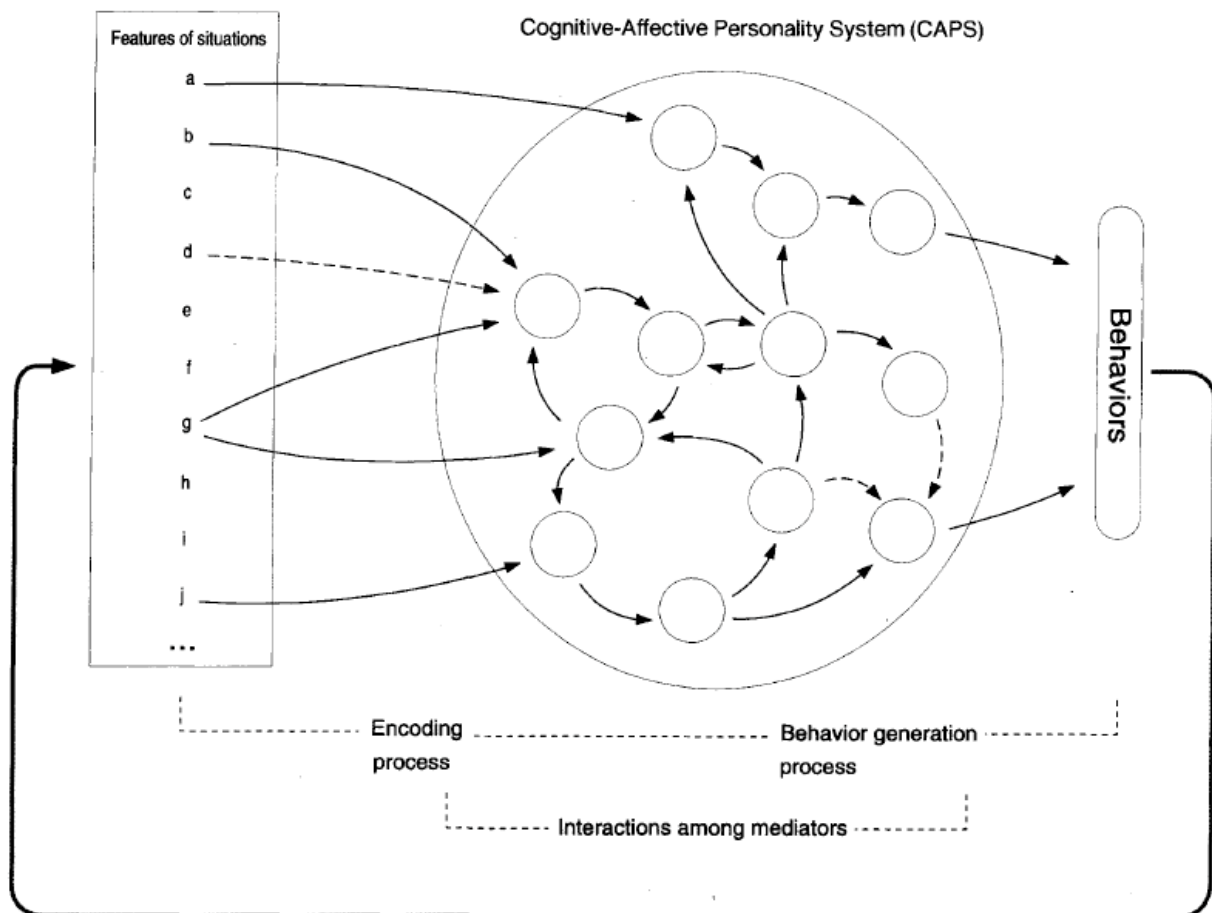
deuxième). La distinction entre structures cognitives encodées de manière stable dans la mémoire et processus cognitifs actifs pendant un laps de temps déterminé n'est pas respectée. Mischel compte parmi les « attentes » (3) aussi bien des structures cognitives stables et chroniquement accessibles (ex : l'attente constante d'être rejeté) que les processus qui génèrent des attentes contingentes (ex : croire qu'on sera rejeté par un partenaire potentiel parce qu'on vient de remarquer des signes négatifs dans son attitude). Les critères d'individuation sont flous, ce qui augmente le risque d'une simple traduction non-informative de vieilles idées dans un jargon cognitiviste²⁵⁸. Comme nous l'avons admis depuis le début, il n'y a aucune garantie préalable qu'une science vraiment informative de la personnalité soit possible : les profils de personnalité pourraient émerger de façon chaotique de l'action de milliers de facteurs « moléculaires », sans aucune généralisation « molaire » à découvrir. C'est une vraie possibilité empirique. Mais d'un point de vue méthodologique il faut rechercher autant de précision que possible.

Aspects communs des approches socio-cognitives

Les théories socio-cognitives contemporaines s'accordent sur des principes généraux et diffèrent sur des points spécifiques. Cherchons à saisir le socle commun méta-théorique, avant de les passer en revue. Le thème fondamental est la conception de la personnalité comme un phénomène *émergent* à partir de l'interaction d'un ensemble complexe d'unités cognitives avec l'environnement. Cela exclut toute tentative d'explication directe d'une tendance comportementale par la postulation d'une structure « isomorphe » qui en serait responsable : si les gens sont plus ou moins jaloux, égoïstes ou extravertis cela n'est pas expliqué par des circuits ou « modules » de la jalousie, de l'égoïsme ou de l'extraversion. On ne peut pas dériver la structure cognitive de façon aussi directe à partir des manifestations d'un trait. Au contraire, les tendances comportementales émergent de l'activité d'un *réseau connexionniste* qui relie des « unités cognitives » entre elles et aux stimuli environnementaux. Le connexionnisme dont il est question n'est pas celui qui concerne directement les neurones et leurs réseaux : les unités sont des structures cognitives encodées de façon stable dans le cerveau. Il s'agit de croyances, désirs, buts, standards évaluatifs, épisodes précis, représentations de soi-même ou d'autrui, etc. (pratiquement toute structure isolable encodée dans le système cognitif). Mais le principe est le même : ces unités cognitives à un moment donné peuvent être actives et influencer le déroulement de la cognition ou ne pas être actives. Elles sont reliées

²⁵⁸ Cervone développe davantage la critique des approches les plus anciennes, voir Cervone, "The Architecture of Personality," 185.

entre elles par des liens d'activation ou d'inhibition plus ou moins forts. Ainsi, penser à des vacances à la mer active toute une série de souvenirs précis, des images de soi-même et des idéaux (voyager, faire le tour du monde), mais inhibe (par exemple) l'idée de travail et de maladie. Des connexions du même genre existent entre les stimuli externes et les unités cognitives. Par exemple, une image précise ou une odeur peuvent évoquer en moi (mais pas chez autrui) l'idée de vacance à la mer, ce qui à son tour active ou inhibe les unités mentionnées. L'état d'activation du réseau à un moment donné dépend de l'équilibre entre les forces d'activation/inhibition. Il ne faut pas se faire une idée réductrice de la notion d'activation : penser à mes vacances à la mer n'engendre pas un simple souvenir ou une prise de conscience de mon idéal de grand voyageur. Si cet idéal s'active fortement, cela va influencer le reste de mon activité cognitive à tous les niveaux. Mon humeur peut changer, je vais me sentir plus confiant, assumer des styles cognitifs et explicatifs différents et ainsi de suite. Dans le modèle, une activation pour des raisons anodines (un détail que je remarque dans une pièce) peut avoir des conséquences très fortes en principe, si d'autres unités cognitives n'inhibent pas les réactions concernées. En général, la connectivité du réseau ne représente pas de façon fiable la qualité de la mémoire ou le niveau de conscience explicite : un souvenir « neutre » peut avoir très peu de connexions significatives d'un point de vue émotionnel et motivationnel et être pourtant très bien mémorisé et accessible. Dans l'autre sens, l'association est plus régulière : il est plus facile qu'un souvenir très saillant au niveau émotionnel soit retenu et reste consciemment accessible, mais il y a des exceptions. On peut très bien refouler et ne pas être pleinement conscient de structures cognitives qui restent très influentes au niveau comportemental (ex : agir sur l'influence d'un idéal perfectionniste de soi).



Ce modèle connexionniste²⁵⁹ a d'autres propriétés intéressantes. Il y a une différence entre des structures « accessibles de façon chronique » (*chronically accessible*) et d'autres accessibles de façon épisodique. Les premières sont activées très facilement et fréquemment par le système. Quelqu'un obsédé par sa timidité va y penser tout le temps et interprétera presque toute situation à partir de la peur d'avoir été trop timide. Au contraire, quelqu'un peut être en général très confiant, mais perdre toute assurance si on arrive à le mettre dans un état d'esprit associé à des souvenirs douloureux. La personnalité sera plus lourdement influencée par des structures facilement

²⁵⁹ Mischel et Shoda, "A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure," 254. Le réseau envisagé par les théoriciens socio-cognitifs n'est pas simplement feed-forward, car le comportement engendré par le système affecte à son tour la perception sociale. Il y a bien un cercle : par exemple, si je parle en public souvent (comportement), je vais ressentir moins d'anxiété, et la situation va me paraître moins menaçante (perception sociale). À son tour, cette nouvelle perception va me permettre de faire de choses que je n'aurais jamais faites : cela modifiera davantage ma perception, et ainsi de suite.

activables que par les autres. Cette distinction (applicable au niveau de chaque idéal, souvenir ou image de soi) a aussi un équivalent au niveau global de l'état du réseau : il y a des « états d'esprit » (états d'équilibre du réseau d'activations) dans lesquels on se retrouve plus facilement et par défaut et d'autres qui sont très influencés par la situation. Pour être précis, aucune situation n'est littéralement neutre, car il y a toujours une quantité énorme d'activations et d'inhibitions situationnelles : on n'est pas exactement la même personne lorsqu'on est chez soi, au boulot ou dans une fête. Mais certaines de ces situations (pas les mêmes pour tout le monde) changent de façon moins significative notre état d'équilibre par défaut. Il y a donc des profils d'activation chroniques et des profils situationnels.

Quelle est l'explication du comportement avancée par ce modèle ? Par exemple, comment explique-t-on une conduite courageuse au travail ? Chez quelqu'un qui est régulièrement courageux au travail la connectivité du réseau est telle qu'elle garantit, par rapport aux situations typiquement rencontrées, la stabilité des dispositions émotionnelles nécessaires pour agir de façon courageuse. Cela n'arrive pas facilement qu'un événement inattendu active une peur exagérée ou un manque de confiance. Dans le vocabulaire du connexionisme, la stabilisation du profil d'activation est une forme d'équilibre du système appelé « attracteur ». Suite à un stimulus très saillant, le niveau d'activation des unités varie très rapidement jusqu'à atteindre un niveau d'équilibre où il n'y a que de petites variations (ou en tout cas des variations qui ne sont pas significatives). Tant qu'il n'y aura pas d'autre stimulus capable de menacer l'équilibre (stimulus extérieur ou événement mental intérieur, ex. : commencer à réfléchir à des événements passés douloureux) la liste des unités particulièrement actives et de celles qui influencent très peu la cognition ne change pas. Un profil d'activation correspond à un état d'esprit global qui favorise certaines actions ou réactions émotionnelles. Dans le modèle, la cohérence du comportement dans un ensemble de situations s'explique par la stabilité du profil d'activation correspondant : la différence entre trois personnes qui sont respectivement honnêtes uniquement au travail, uniquement dans leur vie familiale ou dans les deux contextes est que pour la première les situations typiques qu'on rencontre dans la vie professionnelle poussent le système vers un attracteur qui facilite des réactions honnêtes ; pour la deuxième ce sont les événements typiques de la vie familiale qui génèrent un tel effet, alors que chez la troisième l'activité du réseau est attirée par l'attracteur de l'honnêteté dans les deux contextes.

Il faut souligner que tout état d'activation possible du système (la liste du niveau d'activation de chaque nœud à un moment t) n'est pas l'un des attracteurs du système car il ne correspond pas à une condition d'équilibre. Des conditions extrêmes peuvent me pousser dans un état d'esprit de colère : je pourrais insulter la personne concernée ou la frapper avec le premier objet à portée de main. Mais

cela ne dure que quelques secondes et il ne s'agit pas d'un attracteur du système. En ce sens, chaque système de personnalité a un nombre fini d'attracteurs, quoique vraisemblablement vaste.

Ainsi, le niveau de cohérence dans la manifestation d'un trait est représenté de façon très simple et élégante dans le modèle : chaque trait est un attracteur dont la cohérence correspond aux dimensions du bassin d'attraction. Si je suis honnête dans un nombre extrêmement vaste de situations le bassin de mon attracteur honnêteté est très vaste, alors que si je suis honnête qu'au travail il est beaucoup plus restreint. Le profil de personnalité d'une personne correspond (traits larges ou étroits) à la liste de ses attracteurs (bassin large ou restreint). Enfin, tout comme la mémoire est vraisemblablement codée par la distribution des forces d'activations et inhibition dans l'ensemble des connexions synaptiques, dans les modèles socio-cognitifs, un profil de personnalité est représenté par la stabilité des liens d'activation/inhibition entre les nœuds. La force des liens d'activation ne peut pas changer trop rapidement au fil du temps : c'est un aspect important du modèle, car autrement il n'y aurait aucune forme de stabilité dans la personnalité. Si le poids des connexions pouvait changer facilement, un timide pourrait devenir extraverti au bout de quelques mois. Mais nous avons vu que la stabilité temporelle des réactions émotionnelles et comportementales a été prouvée de façon plutôt convaincante. Néanmoins, la possibilité de changer le poids des connexions est la question la plus importante par rapport aux conséquences pour l'éthique de la vertu. Nous verrons ci-dessous que, dans les limites d'une prise en compte réaliste des données sur la stabilité temporelle, les approches socio-cognitives sont modérément optimistes quant à la possibilité du changement. C'est l'aspect le plus important pour le dialogue avec l'éthique de la vertu.

Questions ouvertes : la nature des unités cognitives et le rôle des modules dans la cognition sociale

Dans le cadre connexionniste général, les variables qui différencient les modèles spécifiques sont une attitude oecuménique ou sélective par rapport (1) aux systèmes cognitifs postulés et (2) au type de structures cognitives admises.

En ce qui concerne ces dernières, on peut accepter que toute croyance, désir, but, souvenir ou idéal régulateur représente un nœud du réseau ou restreindre la liste possible : peut-être les souvenirs n'influencent-ils pas le profil de personnalité de façon significative, alors que les idéaux régulateurs et l'activation du module de gestion de la réputation ont un rôle majeur. L'enjeu théorique est le contenu empirique du modèle : une théorie qui limite la liste des facteurs importants prend des

risques, car elle est plus facile à modéliser, fait des prévisions plus précises et facilement falsifiables. Par contre, en démultipliant les types de nœuds possibles on se donne un certain avantage, mais on perd du pouvoir explicatif.

En ce qui concerne les systèmes cognitifs ou neurobiologiques, les théoriciens socio-cognitifs cherchent à en postuler le moins possible. La tendance est de rejeter l'interprétation modérée du connexionnisme, selon laquelle une architecture en réseau à un niveau d'analyse est compatible avec une architecture cognitive classique à d'autres niveaux d'analyse. Si on n'est pas connexionniste, il est inévitable de concevoir l'esprit comme un ensemble de dispositifs cognitifs en interaction. Ces dispositifs peuvent être plus au moins « modulaires ». Ils sont d'autant plus modulaires qu'ils traitent des inputs spécifiques et qu'ils se rapprochent d'une « boîte noire » dans leur mode de fonctionnement : le traitement de l'information entre l'input reçu et l'output engendré reste inaccessible à d'autres systèmes cognitifs (par exemple, le traitement de l'information visuelle). La différence en modularité entre plusieurs systèmes cognitifs est sans doute une question de degré, mais, en général, on n'est pas obligé de considérer l'architecture standard « systèmes cognitifs (plus ou moins) modulaires en interaction » comme une *alternative* au connexionnisme. En effet, (1) chaque module pourrait être « réalisé » par un réseau connexionniste (à l'intérieur de la boîte noire, il y aurait un réseau connexionniste qui traite l'information) et (2) les modules/systèmes pourraient « dialoguer » de façon connexionniste (s'activer ou s'inhiber l'un avec l'autre comme les nœuds d'un réseau).

Selon l'interprétation la plus radicale du connexionnisme, l'architecture en réseau est bien une alternative à une architecture classique : il n'y aurait pas de systèmes ni de modules, *parce qu'*il y a des réseaux. L'interprétation radicale est considérée comme le véritable fond de commerce du connexionnisme.

Pourtant, dans le schéma méta-théorique de l'approche socio-cognitive, un système cognitif pourrait bien coïncider avec un nœud du réseau (hypothèse du « dialogue » connexionniste). Dans ce cas de figure, ce serait l'activité même du système qui est mise en marche ou inhibée par le poids des connexions. Supposons qu'il existe un module consacré à la gestion des menaces à ma réputation : dans une situation sociale où je risque d'être ridiculisé, le module s'active et m'alerte. Dans une interprétation connexionniste, certains stimuli environnementaux ou l'évocation d'un mauvais souvenir peuvent altérer (stimuler ou inhiber) le niveau d'activité du module lui-même : à partir de ce moment-là, ma cognition sera (par exemple) influencée par la peur de ne pas être à la hauteur et d'être ridiculisé.

Le débat interne sur le rôle des systèmes cognitifs est épineux, car de cette littérature ressort de façon plutôt évidente une aversion (souvent déclarée)²⁶⁰ aux architectures cognitives très modulaires, notamment dans leur version évolutionniste. Il y a des raisons aussi bien stratégiques que théoriques pour cela, mais il serait difficile de ne pas y voir une certaine résistance idéologique. L'intérêt stratégique à limiter la prolifération des modules est celui qu'on vient d'évoquer : la spécificité d'une explication connexionniste « radicale » de la stabilité de la personnalité serait perdue si on introduisait une liste trop longue de modules très influents (cela ressemblerait à une architecture cognitive classique formée par plusieurs systèmes qui dialoguent entre eux).

En même temps, une économie trop stricte en termes de modules rend la théorie peu plausible : il est difficile d'imaginer que le nombre considérable de régularités qu'on retrouve dans la cognition sociale soit produit par le simple jeu d'activation/inhibition de liens largement contingents entre croyances, désirs, stéréotypes et d'autres représentations.

On peut revenir à l'exemple cité ci-dessus : la gestion de la réputation. La psychologie sociale a montré de façon assez claire que le « schéma du soi » (*self-schema*) a une influence très puissante sur la cognition. Dans toute situation sociale, il y a une image précise de nous-mêmes que nous voulons faire passer. D'ailleurs, ce n'est pas toujours la même : on retrouve la distinction entre des aspects chroniques et situationnels. Le niveau de généralité des éléments qui composent cette image peut varier : il peut s'agir de « paraître sûr de soi » (en général) ou « capable de résoudre un problème informatique ». Si un facteur est très saillant dans une situation, on ira adopter des stratégies pour le préserver, et cela de façon tout à fait automatique et non consciente. Par exemple, dans un contexte où la réputation d'expert en informatique est très saillante, ce sera facile de sous-évaluer la difficulté d'un problème technique, car admettre à soi-même que le problème est complexe est coûteux : on risque de mettre en danger sa propre image d'expert. Les biais cognitifs générés par le fait d'avoir une réputation à défendre peuvent avoir des effets dévastateurs et pousser les gens à des ré-interprétations systématiques de la réalité. Cela est probablement à la base de la formation des idéologies élitistes : dans un groupe, le fait d'avoir un privilège ou un avantage (qu'il soit mérité ou non) à un moment donné (être plus riche, habiter dans un meilleur quartier, avoir passé un examen d'entrer dans une école d'élite très difficile) rend très saillante une réputation basée sur des qualités supérieures à la moyenne. Ensuite, sans aucune intention consciente tout est ré-interprété de manière à préserver cette réputation. Le problème n'est pas uniquement le maintien

²⁶⁰ Voir Cervone, "PERSONALITY ARCHITECTURE: Within-Person Structures and Processes," 433.

de la discrimination, mais aussi la chute du niveau qualitatif de l'élite, dès lors que ses résultats sont déconnectés d'une évaluation objective (celle-ci constituerait aussi une menace)²⁶¹.

Rendre compte de phénomènes de ce genre uniquement par un réseau de liens de renforcement/inhibition entre des structures cognitives n'est pas très convaincant. Il semble que les processus cognitifs soient dirigés dans un sens très précis et il est difficile de ne pas postuler l'existence de quelques systèmes cognitifs spécifiques consacrés à la cognition sociale. Pour les théoriciens socio-cognitifs, l'enjeu est donc celui de limiter le plus possible la modularité de l'architecture proposée, tout en préservant la plausibilité du modèle.

Il y a aussi des raisons théoriques : les théoriciens socio-cognitifs pensent que la place naturelle des systèmes cognitifs spécifiques dans l'architecture de la personnalité est celle des systèmes neuro-cognitifs responsables de tendances générales et peu fiables. Dans le chapitre sur la théorie des traits, nous avons défendu une interprétation réaliste et causale de traits généraux tels que l'extraversion ou le névrotisme. L'existence de ce type de traits, qui n'influencent le comportement qu'au niveau de tendance moyenne, a une explication évolutive plausible. La variation observée dans le comportement animal est expliquée par des traits homologues de ceux qu'on retrouve chez l'homme. La variation est probablement due au niveau d'activation différent de systèmes neurobiologiques dont l'étude est bien avancée. Par exemple, certaines personnes ont besoin d'un niveau de stimulation plus élevé et ont tendance à pratiquer des sports extrêmes, à voyager ou chercher l'aventure²⁶².

Les théoriciens socio-cognitifs acceptent l'idée d'une architecture de la personnalité à deux niveaux : des systèmes neurobiologiques responsables des traits généraux (c'est le tempérament, ex : avoir besoin d'aventure) et un réseau connexionniste cognitivement plus sophistiqué qui est responsable des réactions dans des situations précises (dans quels contextes précis on cherche l'aventure et dans quels contextes on est plus calme). Il s'agit de la distinction entre tempérament et personnalité proprement dite²⁶³. Le tempérament concerne des systèmes dont l'activité est étroitement liée aux émotions, alors que les processus cognitifs au niveau de la personnalité ne sont pas nécessairement définis par l'activation émotionnelle. Mischel a utilisé la métaphore de la

²⁶¹ Pour une série d'analyses consacrées à cet effet, voir le numéro spécial *Social justice research* 17, no. 2 (2004).

²⁶² Il s'agit de la « recherche de sensations » (*sensation seeking*), étudiée depuis des décennies par Zuckerman, voir en particulier le chapitre 5 de Zuckerman, *Psychobiology of personality*. L'ouvrage dans sa globalité offre une introduction excellente à la neurologie et à la psychobiologie de la personnalité.

²⁶³ Cloninger, "8 Completing the Psychobiological Architecture of Human Personality Development: Temperament, Character, and Coherence."

cognition « chaude » et « froide » et d'un système « cognitif-affectif »²⁶⁴. Si on veut être rigoureux, les systèmes tempéramentaux ne sont pas à l'extérieur du réseau connexionniste : ils peuvent être catégorisés comme des nœuds, car un stimulus pertinent peut les activer et engendrer les conséquences prédites par la théorie (ex : une surexcitation et un besoin d'aventure). Ils forment néanmoins une partie très spécifique et sui generis du réseau, et c'est exactement dans cette région que les psychologues socio-cognitifs voudraient cantonner la présence des modules. Au premier abord, cela semble être cohérent aussi d'un point de vue évolutionnaire. D'un côté, les arguments les plus puissants en faveur d'une architecture modulaire sont de type évolutionnaire, de l'autre le tempérament est l'ensemble de systèmes cognitifs dont la présence d'homologues chez les animaux est la plus claire. Cela se tient et on peut y voir une convergence théorique : dans le domaine de la personnalité, l'héritage évolutif principal est le macro-système modulaire du tempérament. Malgré tout cela, il serait difficile de nier la présence d'une attitude idéologique négative envers la psychologie évolutionniste et la modularité de l'esprit chez ces théoriciens. Dans l'une des meilleures revues récentes de la littérature socio-cognitive, (à ma connaissance la seule qui affronte directement la question), Cervone accuse les approches évolutionnistes d'ignorer les interactions gène-environnement :

« (...)Heyes passe en revue les données relatives à quatre processus de développement possibles pour les systèmes cognitifs, parmi ceux-ci il y en a qui sont caractérisés par des interactions gène-environnement, et les modèles de la psychologie évolutionnaire innéiste traditionnelle ne rendent pas compte de cela »²⁶⁵.

Or, il est clair qu'une approche évolutionniste raffinée admet sans problèmes la coexistence de plusieurs stratégies issues de l'évolution mais conçues pour être activées dans des conditions environnementales et sociales différentes, et aussi la possibilité de développer des traits partiellement ou complètement différents suivant les conditions rencontrées pendant le développement de l'organisme. Il s'agit du phénomène de la plasticité phénotypique : un génotype est dit plastique s'il acquiert une morphologie différente selon l'environnement. Le génotype non plastique produit le même phénotype quel que soit l'environnement. Par exemple, l'abeille devient une travailleuse ou une reine suivant la quantité et qualité de l'apport nutritif reçu pendant la phase de larvaire²⁶⁶. L'épaisseur de la coquille des littorines du golfe du Maine dépend de la présence ou absence d'une espèce déterminée de crabes prédateurs : à partir du même génotype, la littorine

²⁶⁴ Mischel et Shoda, "A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure."

²⁶⁵ Cervone, "PERSONALITY ARCHITECTURE: Within-Person Structures and Processes," 433.

²⁶⁶ Mameli et Bateson, "Innateness and the Sciences."

développe une coquille plus épaisse en présence des crabes²⁶⁷. Il n'y a pas de raisons pour que l'approche évolutionniste à la cognition animale et humaine se prive de ces procédés classiques de la théorie néo-darwinienne. Les théoriciens socio-cognitifs parlent peu de l'évolution et ne semblent pas prendre en considération les approches les plus raffinées. Leur attribuer un certain préjugé me semble justifié. Peu importe : leur programme de recherche demeure la seule tentative d'étude scientifique de la structure cognitive de la personnalité dans l'état actuel de nos connaissances, il faudra en accepter les résultats et repérer les tendances qui s'en dégagent. Passons en revue les modèles principaux et les résultats empiriques obtenus, pour chercher à comprendre ce qui reste de l'idéal de construction d'une personnalité cohérente et solide, et quelles sont les conséquences pour une théorie de la vertu contemporaine.

KAPA (knowledge appraisal personality architecture)

La contribution de Cervone relève surtout de la clarification du cadre conceptuel²⁶⁸. Ayant accepté la distinction entre tempérament (émotionnel et ancien dans la phylogenèse) et personnalité proprement dite (plus évoluée d'un point de vue cognitif, récente dans l'évolution et en charge de la gestion du premier système), ses efforts se concentrent sur la deuxième. L'apport fondamental de sa théorie est la distinction entre structures de connaissance codées dans le système (*knowledge*) et processus de jugement et interprétation (*appraisal*). Une structure de connaissance est un état mental doté d'un contenu représentationnel codé et stocké dans le système : d'après Cervone, les trois types d'états mentaux les plus importants pour expliquer la cohérence du comportement sont les croyances, les désirs et les normes évaluatives (*evaluative standards*). Ces états sont les unités ou nœuds du réseau connexionniste. Par contre, un jugement est le résultat d'un processus de traitement de l'information et ne correspond pas à une connaissance déjà stockée. Je peux savoir que j'aime les personnes stimulantes qui me mettent à l'épreuve (connaissance), et juger que telle personne va certainement attirer mon attention pendant une soirée (jugement). La différence ne recoupe ni celle entre « chroniquement accessible » et « situationnel », ni celle entre « général » et « spécifique ». Une représentation idéale de moi en tant qu'athlète performant peut être très peu accessible (elle est activée uniquement dans des situations spécifiques) mais clairement stockée

²⁶⁷ David et Samadi, *La théorie de l'évolution* 203-208. L'évolution favorise la plasticité phénotypique lorsque l'adoption totale d'une stratégie alternative chez un individu serait trop coûteuse : le désavantage d'une littorine qui développait *toujours* une coquille très épaisse – même en absence de crabes – serait trop grande, à cause du coût métabolique très important de la coquille.

²⁶⁸ Pour la meilleure présentation de son approche, voir Cervone, "The Architecture of Personality."

parmi mes standards évaluatifs. De la même façon, une représentation peut être très spécifique (« je suis un expert de cocktails ») mais cruciale dans l'économie de ma personnalité.

Un jugement est généré par le système cognitif dans une situation déterminée et n'était pas stocké dans la mémoire auparavant. Si un soir j'arrive à la conclusion que je suis un expert de cocktails parce que j'ai un peu d'expérience dans le domaine, mais surtout parce que je suis très confiant en moi, il s'agira d'un jugement. Si je sais que je suis expert de cocktails et que cela joue un rôle dans mon schéma du soi (*self-schema*), il s'agira d'une structure de connaissance. La différence est intéressante pour deux raisons. D'abord, la clarté conceptuelle : en psychologie, il est important de savoir quelle sorte d'entités et processus sont postulés. On n'a pas le droit d'être vague sur la liste de représentations stockées dans le système. Autrement, on renonce à la testabilité indépendante de l'existence et des effets de ces représentations. Ensuite, les jugements n'auront pas les mêmes effets que les représentations dans les dynamiques du système de personnalité, car ils ne constituent pas des nœuds du réseau.

Dans notre classification des théories socio-cognitives, KAPA est plutôt parcimonieuse en termes d'unités cognitives admises : Cervone pense que les désirs, les croyances et les standards évaluatifs jouent le rôle principal. Il n'est pas aussi libéral que Mischel, qui admet des états aussi spécifiques que des souvenirs déterminés. De la même façon, il postule l'existence de très peu de systèmes cognitifs: l'ensemble des structures qui forment le soi cognitif (croyances sur soi, standards évaluatifs) est central mais ne joue pas un rôle exclusif, comme dans la théorie de Morf (voir ci-dessous). Cervone pense que des standards éthiques et d'autres standards évaluatifs généraux qui ne se réfèrent pas au comportement du sujet (autres que : soi-même en tant que parent, soi-même en tant que professeur etc.) sont également décisifs.

Modèle de l'auto-régulation de Morf

Les théoriciens socio-cognitifs partagent une vision de l'architecture de la personnalité à deux niveaux. La fonction du système plus récent dans la phylogenèse est celle de réguler le système plus ancien et émotionnellement chargé. Tous s'accordent sur le fait qu'une partie fondamentale de cette régulation consiste dans la stabilisation et gestion du « soi cognitif ». De manière générale, par soi cognitif dans la littérature contemporaine en psychologie de la personnalité, on entend l'ensemble des processus qui stabilisent, défendent, négocient et changent une série de représentations de soi

même²⁶⁹. Il y a deux conceptions du soi : l'une descriptive et l'autre idéale. Dans les deux cas, les représentations peuvent être générales ou spécifiques (on croit être un bon parent, on rêve d'être un excellent joueur de poker) et jouent un rôle décisif dans la vie sociale et l'explication du comportement. Les psychologues estiment qu'une grande partie de nos actions visent à défendre une certaine image de soi-même, à la négocier et à la diffuser²⁷⁰.

Dans le cadre de ce consensus, les théoriciens de la personnalité se partagent sur deux points. Premièrement, il y a ceux qui pensent que la notion de « soi cognitif » se réfère à un ensemble important mais hétérogène de processus cognitifs : la seule option serait d'étudier séparément chacun de ces processus. D'autres considèrent qu'il faut postuler un véritable système cognitif du soi qui fonctionne de manière cohérente et coordonnée. Deuxièmement, il y a ceux qui pensent que le soi-cognitif n'est qu'une partie importante de la personnalité – mais une parmi d'autres – et ceux qui croient que ses dynamiques expliquent la plupart des régularités du comportement. Morf pense qu'il faut postuler un vrai système cognitif du soi et que son activité est la plus importante dans le réseau connexionniste responsable de la stabilité de la personnalité. L'architecture cognitive de la personnalité serait constituée par des processus d'auto-régulation. Le soi-cognitif serait composé par quatre sous-systèmes :

Conceptions de soi : l'ensemble des représentations de soi (descriptives ou idéales) ainsi que des sentiments d'estime et de respect de soi. L'objectif global des dynamiques à l'œuvre dans le soi cognitif est de préserver une configuration optimale de ces représentations et évaluations affectives.

Processus intra-personnels : l'ensemble des processus cognitifs, évaluatifs et émotionnels qui influencent la sélection et l'interprétation de l'information pertinente pour le soi. Par exemple : l'interprétation du succès ou de l'échec, la mémoire sélective d'épisodes passés ou l'interprétation biaisée du feedback qu'on reçoit dans les interactions sociales.

Processus inter-personnels : il s'agit de processus au cours desquels le sujet modifie (consciemment ou inconsciemment) son environnement social pour affecter l'image de soi. Par exemple : éviter des situations où l'on pourrait recevoir du feedback négatif par rapport à l'image de soi comme un grand danseur, ou créer des contextes où l'on est sûr d'avoir le soutien et les confirmations recherchés.

Relations sociales significatives : comme nous l'avons déjà vu en discutant les schémas relationnels (voir chapitre sur la théorie des traits du sens commun), le soi-cognitif est défini par un ensemble de

²⁶⁹ H. Markus et E. Wurf, "The Dynamic Self-Concept: A Social Psychological Perspective," *Annual Review of Psychology* 38, no. 1 (1987).

²⁷⁰ Voir W. Mischel et C. C. Morf, "The self as a psycho-social dynamic processing system: A meta-perspective on a century of the self in psychology," *Handbook of self and identity* (2003).

relations particulièrement significatives, normalement avec des proches (famille, parents, compagnon, ami) ou des personnes qui ont beaucoup marqué notre vie. Une relation significative a plusieurs effets majeurs sur le soi cognitif. D'abord, la conception descriptive et la conception idéale de soi-même peuvent être influencées par celles de l'autre ou par la conception que l'autre a du sujet (ex. : un idéal de moi-même en tant que bon athlète peut être aujourd'hui chroniquement accessible parce que mon père était un bon athlète ou parce que il voulait que je devienne un bon athlète). Ensuite, même si une relation n'a pas affecté la conception de moi-même mobilisée par défaut, lorsqu'un stimulus pertinent active le schéma relationnel (une chanson me fait souvenir d'un ami proche que je ne vois pas depuis longtemps), les conceptions de moi-même liées à cette personne (ce à quoi j'aspire lorsque je suis avec elle, ses aspirations pour moi) s'activent et commencent à influencer mon comportement.

La thèse fondamentale de Morf est la suivante : la cohérence (les régularités qui peuvent être mises en évidence au niveau des réactions émotionnelles et comportementales) de la personnalité est expliquée par la configuration et les différentes typologies de problèmes du soi-cognitif. Cela paraît crédible si on se base sur les types de personnalités socio-cognitifs le plus étudiés pour l'instant, notamment la sensibilité au rejet, le narcissisme et la personnalité dépendante. Il s'agit d'autant de cas où le sujet déploie des stratégies complexes pour défendre les points faibles de sa conception du soi : un manque de valeur, un manque d'autonomie et une faible estime de soi. Néanmoins, le choix des cas d'études est évidemment biaisé en faveur de la théorie. Le but à long terme de l'approche socio-cognitive est celui de développer une typologie riche qui explique aussi bien les aspects positifs de la personnalité humaine que les troubles. Cette entreprise n'en est qu'à ses balbutiements et il est prématuré d'évaluer le rôle joué par le soi-cognitif dans l'architecture de la personnalité.

CAPS (cognitive-affective personality system)

Le modèle CAPS développé récemment par Mischel et Shoda est celui qui représente le mieux les caractéristiques générales de l'approche socio-cognitive. Il est caractérisé par une attitude très œcuménique quant aux types de nœuds admis dans les réseaux, parcimonieuse quant au nombre de modules postulés et centrée sur une distinction nette entre éléments affectifs-émotionnels et non-affectifs. En ce qui concerne le premier point, Mischel propose une liste des types de nœuds – qu'il appelle unités cognitives-affectives (*CAU, cognitive affective units*) – « ...qui inclut la représentation et reconstruction du soi, d'autres personnes, de situations, de buts bien enracinés,

d'attentes, d'états émotionnels aussi bien que de souvenirs de personnes et événements passés »²⁷¹.

Presque tout y est admis. Par contre, Mischel, qui est un pionnier et un expert de la psychologie de l'autorégulation, partage le scepticisme de ses collègues quant à la modularité massive de l'esprit et se concentre sur un dualisme fondamental entre un système émotionnel tempéramental et les fonctions d'autorégulation nécessaires pour le gérer, d'où le nom de système cognitif-affectif.

Quels sont les résultats empiriques de ce programme de recherche ? Si le but est l'étude de l'architecture cognitive de la personnalité à l'échelle individuelle ainsi que la modélisation et la prédiction (autant que possible) du comportement d'un individu particulier, on en est encore loin. Pour l'instant, il s'agit de faire en sorte que le modèle prédise (1) le genre de cohérence associée aux traits généraux (le fait d'être en moyenne agréable ou extroverti) et (2) un ensemble de régularités intéressantes et vérifiables expérimentalement. (1) relève de ce qu'en anglais on appelle « *data fitting* » : la capacité d'une théorie de prédire des résultats déjà connus. En effet, comme nous l'avons vu, des décennies d'expériences en psychologie de la personnalité ont prouvé que le niveau moyen des traits généraux est plutôt stable au fil des années. Dans des simulations informatiques développées par Mischel et Shoda, leur modèle semble répliquer ce résultat²⁷². La cohérence dans la manifestation des traits (réactions émotionnelles *et* comportement) est représentée de façon simple et élégante par la liste des attracteurs et l'ampleur de leurs bassins. Un attracteur large correspond à un trait large : un équilibre d'activation qui favorise la timidité sera un attracteur très important pour quelqu'un qui est en général très timide ; quelqu'un qui a uniquement peur de parler en public sera poussé vers l'attracteur timidité dans un nombre inférieur de situations très spécifiques. Il s'agit donc de vérifier l'existence de bassins d'attraction larges (qui correspondent à des tendances par défaut) dans une simulation informatique plausible.

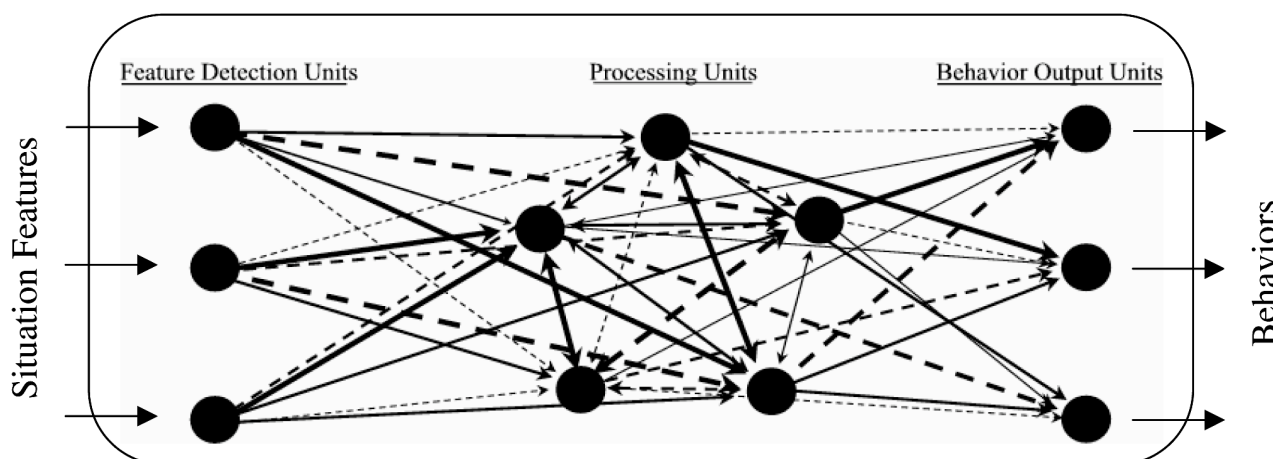
Shoda a simulé une personnalité hypothétique dans un réseau connexionniste de 20 unités : 5 détecteurs de caractéristiques de la situation sociale, 10 unités cognitives-affectives internes et 5 unités correspondant à des réponses comportementales. La nécessité des détecteurs se justifie par l'impossibilité d'une interaction directe entre des situations sociales objectives et externes et les unités internes du système. Le fait d'entrer dans un amphithéâtre ne peut déclencher aucune réaction émotionnelle sans que la situation ne soit « interprétée » par le système cognitif. Dans la simulation, une situation externe affecte les unités internes par le biais du profil d'activation des 5 unités de détection. La fonction de mise à jour des niveaux d'activation des unités était la suivante :

²⁷¹ Walter Mischel, "Toward An Integrative Science Of The Person," *Annual Review* 55, no. 1 (2004): 11.

²⁷² Yuichi Shoda, Scott LeeTiernan et Walter Mischel, "Personality as a Dynamical System: Emergence of Stability and Distinctiveness from Intra- and Interpersonal Interactions," *Personality & Social Psychology Review* 6, no. 4 (2002).

pour chaque unité, au 98% du niveau d'activation du cycle précédent va être ajoutée la somme des inputs reçus. S'il n'y a aucun input, l'activité de l'unité s'affaiblit lentement. Dans ces conditions, le réseau était testé sur 100 situations, correspondant à autant de profils d'activation des unités de détections. Le niveau de chaque unité d'activation varie sur une échelle qui va de -1 à 1 : une situation possible peut être par exemple « 0,3 ; -0,9 ; 0,1 ; -0,2 ; 0,8 », où cette séquence représente l'interprétation « perceptive » de la situation par le système. L'effet de l'exposition à 100 situations a été testé pour 40 « personnalités » différentes (c'est-à-dire pour 40 distributions de l'intensité des liens d'activation/inhibition entre les différentes unités du système). Le résultat a été celui de montrer que chaque personnalité avait un nombre variable d'attracteurs de dimensions différentes²⁷³. Notamment, ces « personnalités » ont toujours quelques attracteurs au bassin large.

Personality System



Ce modèle²⁷⁴ n'est qu'une première tentative, il n'y a pas de prétention à un niveau significatif de réalisme psychologique (il suffit de remarquer qu'il n'y a qu'un nombre total de 20 unités dans le réseau). Cependant, le résultat est intéressant, car le modèle n'a pas de présupposés discutables ou *ad hoc*. Il semble que l'existence de bassins d'attraction importants soit une conséquence directe de l'approche socio-cognitive. En d'autres termes, si l'architecture de la personnalité est formée par un réseau connexionniste, il y aura forcément des tendances moyennes générales dans le comportement généré : par exemple, une certaine timidité de base. L'existence de ces tendances étant l'un des résultats les plus solides en psychologie de la personnalité, cela représente un avantage considérable pour le modèle.

²⁷³ Ibid.: 321.

²⁷⁴ Version simplifiée du modèle (notamment, il y a moins d'unités) dans la figure Ibid.: 318.

Les formes de cohérence « substantielle » proposées par l'approche socio-cognitive : profil de personnalité, types, dynamiques relationnelles et la personnalité d'une culture

Prédire l'existence de traits généraux et moyens est une contrainte d'adéquation de la théorie, mais ce n'est pas ce qu'on attend le plus. Le vrai résultat intéressant est l'exploration d'autres formes de cohérence possibles et prédites par le modèle. Des traits généraux et peu fiables ne peuvent pas être à la base d'une théorie normative ambitieuse du caractère. Certes, on pourrait se contenter d'être en moyenne plus honnête, généreux ou sage, mais il n'y a pas de raisons de renoncer aux ambitions de la théorie de la vertu classique avant de les avoir mises à l'épreuve sérieusement. Évaluer les ambitions pratiques ou théoriques de la façon la plus rigoureuse possible avant de les modérer ou de les rejeter est l'approche méthodologique de cette thèse. L'originalité des approches socio-cognitives tient au fait qu'elles représentent la seule tentative explicite de conceptualiser de manière rigoureuse (grâce aux moyens théoriques des sciences cognitives) de nouvelles formes de cohérence possibles et souhaitables. Il s'agit en même temps de travaux pionniers et de la seule référence existante pour une théorie de la vertu qui veuille avoir des bases empiriques solides.

Les théoriciens socio-cognitifs étudient le problème de la cohérence du comportement à partir de la notion de méta-cohérence du système de personnalité. Un profil de personnalité individuel est défini par la distribution et la force des liens d'activation/inhibition entre les unités du réseau. Cette forme de connectivité ne peut varier que très lentement au fil de la vie. Aucune forme de cohérence émotionnelle et comportementale « intéressante » n'est assurée par ce fait, car ce qui est stable est seulement un profil d'associations entre situations et effets psychologiques. *Si* on est à une fête, *alors* on parle à tout le monde ; *si* on se retrouve face à notre supérieur, *alors* on ne trouve pas de mots. Le simple fait que le profil de si-alors (*if-then*) soit stable n'assure aucune forme de cohérence substantielle intéressante ou digne d'être étudiée. Il s'agit plutôt d'une méta-cohérence : la stabilité du même pattern de réactions, aussi fragmenté qu'il soit. Cela est déjà remarquable : le profil complet de si-alors représente l'unicité d'une personne. Quelles que soient les structures de personnalité communes à plusieurs personnes, deux individus ne pourront jamais régir de façon parfaitement identique à la même liste de situations spécifiques. Il y aura toujours des facteurs idiosyncrasiques mais pourtant stables et il est utile d'avoir des hypothèses sur l'architecture cognitive responsable du phénomène. Un réseau où la configuration exacte des connexions est spécifique à chaque individu – mais stable – nous donne exactement le genre d'explication recherchée. Le terme introduit par Mischel et Shoda à propos de ces profils individuels est *personality signature*.

A ce point, la question importante est celle de savoir si la forme de méta-cohérence postulée par le modèle est compatible et associée à des formes intéressantes de cohérence substantielle. Les théoriciens socio-cognitifs pensent avoir produit des résultats empiriques ou au moins une reconceptualisation fructueuse à trois niveaux d'abstraction : individuel, relationnel et culturel.

Le niveau individuel est le mieux étudié. A mi-chemin entre les gros bassins d'attractions correspondant aux tendances générales et les réactions fortement idiosyncrasiques, on retrouve des tendances comportementales définies par des attracteurs de dimensions moyennes. Les dynamiques psychologiques qui expliquent ces régularités comportementales ont une certaine généralité : il ne s'agit pas de structures de personnalité spécifiques à un individu (ex : une tendance stable à s'énerver dans une liste idiosyncrasique de situations), mais on les retrouve chez plusieurs personnes. En ce sens, elles relèvent de plein droit de l'étude de la personnalité (contrainte de généralité) en tant qu'opposé à l'étude de l'individualité ou identité personnelle (unique). Après la longue histoire de l'approche typologique à la personnalité, Mischel et Shoda pensent pouvoir finalement étudier de façon rigoureuse la notion de type.

«(...) un type de personnalité désigne les personnes qui partagent la même organisation des relations entre les unités du réseaux qui traitent un ensemble de situations. Les types sont définis en termes de processus cognitifs-affectifs spécifiques qui produisent des profils si- alors de pensées, émotions et comportements visibles caractéristiques²⁷⁵».

Sur cette notion se concentrent tous les espoirs de l'approche socio-cognitive. Une typologie socio-cognitive est censée résoudre tous les problèmes associés aux traits larges : des traits comme le névrotisme et l'extraversion ont un pouvoir de prédiction faible pour un individu déterminé (cohérence inter-situationnelle modeste) et l'effet de leur interaction sur le comportement est encore plus incertain. Au fait que la prédiction des réactions d'un individu serait très faible par hypothèse (en tant que résultante de deux tendances moyennes non fiables) s'ajoute le problème de l'homogénéité locale (voir chapitre sur la théorie des traits en psychologie). Les interactions entre les traits sont étudiées à l'échelle de la population : les valeurs moyennes sur la direction et l'intensité de l'interaction sont compatibles avec des rapports très différents entre névrotisme et extraversion dans des sous-populations ou chez quelques individus. Même si nous avons défendu une interprétation réaliste, il ne faut pas oublier que l'existence de ces traits chez les individus est

²⁷⁵ Mischel, "Toward An Integrative Science Of The Person," 14.

mise en question en ce moment. Plusieurs théoriciens pensent que les traits du *big five* ont uniquement une valeur taxinomique²⁷⁶.

Au contraire, par définition une signature est un trait de personnalité (1) actif à l'échelle individuelle ; (2) qui a des effets fiables (cohérence inter-situationnelle élevée) ; (3) qui interagit de façon précise et testable avec d'autres structures de la psychologie individuelle. Autrement dit, il s'agit d'une notion non-suspecte qui va contribuer à aligner la psychologie de la personnalité sur les standards des sciences cognitives.

Malheureusement, les travaux de Mischel et Shoda pour l'instant ne font pas exception à la tendance historique de la psychologie à se concentrer sur les troubles et la pathologie plutôt que sur le fonctionnement optimal. On ne trouvera pas pour l'instant une classification typologique des vertus ou des qualités du caractère. Même en ce qui concerne les troubles, à ce stade le courant socio-cognitif ne peut esquisser qu'une ébauche de typologie. Une typologie est censée être systématique, alors que pour l'instant la littérature socio-cognitive ne s'est intéressée qu'à un nombre limité de types humains, dont on ne sait pas encore s'ils sont mutuellement exclusifs. Malgré cela, il faut s'intéresser à cette littérature, qui représente la seule tentative sérieuse de modéliser la personnalité humaine, permettant un dialogue et un transfert de connaissance entre psychologie de la personnalité et sciences cognitives.

Les deux types humains les plus étudiés par les théoriciens socio-cognitifs sont le narcissique et le type caractérisé par une grande sensibilité au rejet (« *rejection sensitivity* », RS). Voyons de quel genre de résultats empiriques il s'agit.

La sensibilité au rejet

Le type RS est caractérisé par une hypersensibilité à des signes de rejet de la part d'autrui et par une réaction exagérée suite à de vrais épisodes de rejet. Les deux patterns de comportement se manifestent surtout dans le cas de relations importantes, comme des rapport de couple ou d'amitié. A côté d'une certaine prédisposition génétique, des expériences de rejet de la part des parents pendant l'enfance sont un antécédent causal important. L'étude des phénomènes de ce genre est classique dans l'histoire de la psychologie : ce sera intéressant de souligner ce que l'approche socio-cognitive apporte en termes de (1) précision dans la définition des processus cognitifs et des effets comportementaux ; (2) pouvoir de prédiction ; (3) testabilité indépendante de différents processus

²⁷⁶ Voir D. Borsboom, G. J. Mellenbergh et J. van Heerden, "The theoretical status of latent variables," *Psychological Review* 110, no. 2 (2003).

postulés. Si on peut remonter au moins jusqu'à Freud pour l'idée que des rapports problématiques avec les parents influencent la formation de la personnalité²⁷⁷, une notion très importante dans la littérature récente est celle de Style d'attachement (*Attachment Style*). Dans ses travaux pionniers, Mary Ainsworth a exploré comment des bébés âgés de 8 mois et demi jusqu'à 12 mois réagissaient à la séparation de leur mère dans un environnement non familial²⁷⁸. Dans le paradigme de la « Situation étrange », le bébé entre dans une nouvelle pièce où il va jouer en présence de sa mère et d'un inconnu. Ensuite, on étudie la réaction du bébé à différents niveaux de séparation par rapport à la mère. Elle peut être pleinement investie dans le jeu, jouer en montrant des signes de préoccupation et de distraction ou être absente, laissant le bébé avec l'inconnu. Face à ce comportement incohérent de la mère, certains bébés évitent tout contact avec elle, y compris après son retour. On leur attribue un style d'attachement *évitant*. Ceux qui restent sereins pendant toute l'expérience et – dès que possible – cherchent à rétablir une interaction positive avec la mère ont un style d'attachement *autonome*. Ceux qui cherchent un contact avec la mère après son retour mais en même temps paraissent fâchés et sont difficiles à reconforter ont un style d'attachement *ambivalent*. Cette typologie a eu un énorme succès et a été reformulée pour être directement applicable aux adultes. L'influence des styles de l'enfance sur ceux de l'âge adulte est certainement substantiel, même si le déterminisme de l'enfance a été vigoureusement contesté²⁷⁹. En tout cas, l'appartenance à l'un de ces types pendant une phase de la vie a une longue série de corrélats intéressants. Les enfants autonomes à un an sont plus flexibles, ont une meilleure forme d'autocontrôle et savent mieux gérer les interactions sociales que les autres. Les adultes « évitant » ont des difficultés avec l'engagement dans des relations stables, ils sont incapables de donner un support émotionnel à leur partenaire et sont cyniques quant à l'amour. Les adultes ambivalents sont anxieux de perdre leur partenaire et prêts à se sacrifier pour lui faire plaisir. Ils sont très stressés par des séparations et croient que tomber amoureux est facile mais ne dure pas²⁸⁰.

Les corrélats de la typologie des styles d'attachement sont nombreux et importants, mais leur ensemble est délimité de façon vague, par une famille de caractéristiques plus ou moins prototypiques du profil. Les types ne sont pas non plus définis par des hypothèses sur les structures cognitives qui expliquent les réactions comportementales. Par contre, en ce qui concerne la RS, tout

²⁷⁷ On peut citer aussi les travaux de Honey, Erikson et Sullivan, voir J. Pietrzak, G. Downey et O. Ayduk, "Rejection sensitivity as an interpersonal vulnerability," *Interpersonal cognition* (2005).

²⁷⁸ M. D. S. Ainsworth, *Patterns of Attachment: A Psychological Study of the Strange Situation* (Lawrence Erlbaum Associates, 1978).

²⁷⁹ Mischel, *Introduction to personality : toward an integration*.

²⁸⁰ Voir Ibid. 130.

le processus qui va de l'interprétation des situations sociales jusqu'à des réactions typiques (passant par le traitement cognitif) est défini de façon plus précise, chaque niveau étant indépendamment testable. Commenant par l'interprétation des situations, la sensibilité au rejet est caractérisée par une attention extrême à des signes possibles de rejet, qu'ils soient réels ou non. Les situations ambiguës et les commentaires les plus anodins sont systématiquement interprétés comme des menaces.

La présence de cette forme de perception sociale chez les individus RS a été testée. Dans une étude récente, les participants étaient en train d'admirer des œuvres d'art portant sur des sujets différents, dont l'acceptation par autrui, le rejet ou d'autres thèmes positifs et négatifs, mais élaborés en dehors de tout contexte social ou référence à des situations spécifiques. À un certain moment, ils étaient interrompus par un stimulus extrême, par exemple un bruit très fort. Les êtres humains ont un réflexe de surprise involontaire et incontrôlable face à des stimuli de ce genre : ils clignent des yeux, et l'ampleur du clignotement est liée aux niveaux de surprise et menace perçus. Les sujets RS interrompus pendant l'observation d'une œuvre d'art dont le thème est le rejet ont affiché un réflexe beaucoup plus marqué que celui enregistré dans tous les autres cas de figure. Cela constitue une preuve indépendante d'une perception sociale concentrée sur les signes de rejet chez les types RS. En plus, la liste des types de situations pertinentes pour mesurer la sensibilité au rejet est courte : après un long processus de validation, 18 situations ont été sélectionnées dans le questionnaire. Dans chaque situation, le sujet doit imaginer de demander quelque chose de significatif à une personne importante et évaluer sur deux échelles (de 1 à 6) séparées respectivement la probabilité de faire face à un refus et l'anxiété évoquée par cette possibilité.

En ce qui concerne le traitement interne de l'information, un sujet RS est censé manifester des styles explicatifs cohérents avec son attitude. Des expériences ont montré que les types RS ne considèrent pas des explications alternatives pour des comportements ambigus : si on ne suggère pas d'hypothèses alternatives, ils seront persuadés qu'un ami en retard ou pressé a l'intention explicite de les rejeter. Dans la littérature sur la sensibilité au rejet, on retrouve aussi le débat sur la modularité dans l'architecture de la personnalité. Certains théoriciens ont pris en considération l'hypothèse d'un module dont la suractivité serait responsable de la sensibilité au rejet²⁸¹. Des raisons neurobiologiques et évolutionnaires étayaient l'hypothèse. D'un point de vue neurobiologique, les réactions de rapprochement et de fuite (ou inhibition) – produites respectivement par des stimuli positifs et négatifs – représentent l'axe fondamental du système motivationnel dans le monde animal. On retrouve la même structure en remontant très en arrière

²⁸¹ Notamment Downey et Ayduk.

dans la phylogenèse : deux systèmes neurobiologiques distincts (dont l'étude est aujourd'hui très avancée) seraient responsables du comportement de rapprochement et de fuite : l'enjeu est celui d'activer rapidement des réactions automatiques qui permettent à l'organisme de profiter de l'occasion ou d'affronter le danger. Les deux systèmes engagent des réactions émotionnelles intenses, car le rapprochement, le combat et la fuite requièrent des changements immédiats et profonds dans les systèmes circulatoires et hormonaux (accélération du battement du cœur, sécrétion d'adrénaline), et les émotions sont intimement associées à des changements physiques de ce genre²⁸². La sensibilité au rejet est clairement une forme typique d'inhibition rapide et automatique, mais sa portée est beaucoup trop spécifique pour qu'une référence générale au système neurobiologique d'inhibition soit suffisante. C'est ici que des arguments évolutionnaires entrent en scène : dans le milieu évolutif chaque individu était extrêmement dépendant du groupe, et en particulier des membres du groupe avec qui il entretenait des relations étroites. La déstabilisation des relations sentimentales et amicales était un risque recourant et saillant. C'est justement dans ce genre de cas (problème spécifique, récurrent et important) que les psychologues évolutionnistes postulent l'existence d'un module. Ainsi, l'évolution nous aurait doté d'un système cognitif spécifique pour détecter et gérer le rejet dans des situations sociales importantes. A ce point, la question est celle de savoir si on a besoin de postuler un module spécifique ou si l'architecture connexionniste générale est en mesure d'expliquer les données. Des théoriciens tels que Cervone et Mischel ont la tendance à éviter la « démultiplication des modules », alors que Downey et Ayduk y voient une hypothèse intéressante. Nous n'avons pas vocation ici à développer ce débat : il suffira de remarquer que, d'un point de vue méthodologique, l'approche modulaire augmente le contenu empirique et la testabilité de la théorie. À ce stade très primitif de nos connaissances sur la personnalité, cela constitue bien l'avantage principal de la typologie socio-cognitive²⁸³.

Enfin, l'existence d'un bon nombre de réactions comportementales précises a été démontrée dans la littérature. Avant que le sujet RS ne rencontre une manifestation véritable de rejet, on retrouve deux stratégies typiques de prévention. Les deux relèvent du sacrifice de ses valeurs et priorités dans l'intérêt de la relation. D'un côté, le sujet peut être disposé à prendre des risques et à faire des choses qu'il n'accepterait pas dans des conditions normales. Chez des adolescentes, des niveaux

²⁸² Prinz défend l'idée radicale selon laquelle les émotions ne sont qu'une forme de perception de ces changements dans l'organisme, voir J. J. Prinz, *Gut reactions: a perceptual theory of emotion* (Oxford; New York: Oxford University Press, 2004).

²⁸³ En postulant l'existence d'un module spécifique, on prend un risque théorique clair et on s'expose à une falsification potentielle. Par contre, dans un modèle connexionniste de la personnalité, interpréter un comportement recourant comme un simple effet émergent du réseau est l'option non risquée, par défaut.

élevés de sensibilité au rejet ont été associés à la disponibilité (mesurée par un questionnaire deux ans plus tard) à faire quelque chose de mal pour sauver la relation avec leur petit ami. Des aveux authentiques d'actions condamnables ont été également associés à des niveaux élevés de sensibilité au rejet²⁸⁴. De l'autre côté, le sujet peut simplement s'effacer, rester silencieux et ne pas défendre ses intérêts dans la relation. Cela peut aller de la volonté de cacher les événements passés pour ne pas décevoir jusqu'à une tentative d'appropriation des valeurs de l'autre. Dans ce dernier cas, le sujet ne va jamais développer une identité forte et sera radicalement influencé par les partenaires dans les relations qu'il cherche à sauver à tout prix. Dans les termes de la psychologie sociale cognitive, on dira qu'un ou plusieurs schémas relationnels occupent une place exagérée dans l'économie du schéma du soi²⁸⁵. En général, toutes ces stratégies inconscientes visent à minimiser les possibilités de rejet lorsque l'acceptation est encore possible.

Suite à de véritables épisodes de rejet, le sujet RS manifeste rage et agressivité. Le lien entre le rejet et une réaction hostile a été testé dans des études de *priming*. Dans une étude où les participants devaient prononcer le plus rapidement possible des mots à l'écran, un stimulus (subconscient) lié au rejet augmentait la vitesse de prononciation de mots associés à l'hostilité, uniquement chez des sujets RS. L'étude de journaux intimes de couples a montré que des sentiments de rejet étaient suivis par des sentiments d'hostilité chez des femmes RS, alors que pour les autres il n'y avait pas de corrélation significative entre hostilité et rejet²⁸⁶. Dans des interviews en laboratoire portant sur les conflits dans les relations, des femmes RS ont montré beaucoup plus de signes d'agressivité (aussi bien verbaux que comportementaux) que la moyenne. Un investissement affectif important dans une relation amoureuse prédit des réactions violentes chez des garçons RS²⁸⁷. L'ensemble de ces comportements est à la base d'une prophétie auto-réalisatrice. L'hostilité injustifiée du sujet RS finit par exaspérer son partenaire, jusqu'à produire réellement le rejet tant anticipé. L'exaspération

²⁸⁴ G. Downey et O. Ayduk, "Correlates of rejection sensitivity in college students" (Données non publiées: 2002), V. Purdie et G. Downey, "Rejection Sensitivity and Adolescent Girls' Vulnerability to Relationship-Centered Difficulties," *Child Maltreatment* 5, no. 4 (2000).

²⁸⁵ Un schéma relationnel est une conception de soi (valeurs, idéaux, caractéristiques saillantes) associée à une relation avec une personne significative : par exemple, la conception de moi évoquée par la présence (ou le souvenir) de mon père. M. W. Baldwin, "Relational Schemas," *The Coherence of Personality: Social-Cognitive Bases of Consistency, Variability, and Organization* (1999).

²⁸⁶ O. Ayduk et al., "Does rejection elicit hostility in rejection sensitive women," *Social Cognition* 17, no. 2 (1999). Respectivement, expérience 1 et 2. G. Downey et al., "The self-fulfilling prophecy in close relationships: Rejection sensitivity and rejection by romantic partners," *Journal of personality and social psychology* 75, no. 2 (1998).

²⁸⁷ G. Downey, S. Feldman et O. Ayduk, "Rejection sensitivity and male violence in romantic relationships," *Personal Relationships* 7, no. 1 (2000).

des partenaires a été démontrée dans des études de journaux intimes et dans des vidéos d'interactions des couples enregistrés en laboratoire²⁸⁸.

Non seulement le type RS est défini de façon plus rigoureuse et testable que les traits dans la tradition du big five et les typologies précédentes, il existe aussi des études sur l'interaction entre la sensibilité au rejet et d'autres dynamiques psychologiques. Nous allons analyser les interactions entre sensibilité au rejet et (1) capacité d'auto-régulation, (2) personnalité borderline, (3) dépression.

RS et auto-régulation

Toutes les personnes qui anticipent avec anxiété des épisodes de rejet (ce qui est mesurable par un profil spécifique de réactions physiologiques) ne subissent pas la même quantité de conséquences dévastatrices dans leur vie. Cette remarque a stimulé les psychologues socio-cognitifs à étudier les facteurs psychologiques qui pourraient contrecarrer les effets de la sensibilité au rejet. Comme nous l'avons précisé, il y a un accord significatif sur une architecture de la personnalité à deux niveaux : un système plus ancien dans la phylogenèse, bien défini d'un point de vue neurobiologique et fortement émotionnel, et un système plus récent, de plus haut niveau, moins marqué par des réactions émotionnelles et censé coordonner et contrôler le premier système émotionnel. Cette forme de contrôle n'est pas conçue par Mischel comme une soumission de réactions émotionnelles débridées aux directives d'un système cognitif « froid », selon le modèle de la soumission des émotions à la raison. Du point de vue des psychologues, le critère d'une autorégulation réussie est l'adaptation, c'est-à-dire au moins la normalité ou le succès dans la vie professionnelle (école, université, travail) et les relations sociales (amitié, amour). Ce comportement adapté n'est pas assuré par une compétence générale d'autorégulation mais par un ensemble de techniques et de compétences de régulation spécifiques (*skillset*). La relation entre la composante émotionnelle « chaude » et les processus cognitifs « froids » varie selon les besoins stratégiques : par exemple, pour résister à la tentation, il faut savoir éloigner l'attention des émotions qui nous poussent à céder, mais pour persévérer dans un tâche difficile, il faut savoir raviver le désir²⁸⁹.

Mischel est célèbre pour le paradigme expérimental du délai de gratification. Des enfants d'âges différents doivent choisir entre avoir un bonbon tout de suite ou deux bonbons après un délai d'attente d'une vingtaine de minutes. Mischel a conduit, filmé et analysé des expériences de ce

²⁸⁸ Downey et al., "The self-fulfilling prophecy in close relationships: Rejection sensitivity and rejection by romantic partners."

²⁸⁹ Sans arriver au point où le désir deviendrait trop fort et imposerait une satisfaction immédiate.

genre depuis presque une quarantaine d'années. Parmi les variables les plus intéressantes, on retrouve : la présence ou absence d'un expérimentateur, la présence ou absence du bonbon immédiatement disponible et des deux bonbons pour lesquels il faut attendre, la substitution des bonbons par des images de bonbons pendant la période d'attente (après son choix, l'enfant pouvait avoir un ou deux bonbons réels). Cette tradition de recherche montre le pluralisme stratégique de l'autorégulation. S'il ne possède pas la capacité de se distraire de son désir, l'enfant ne réussit pas à attendre. Mais si les bonbons promis ne sont pas face à lui, l'incapacité à raviver le désir avec l'imagination peut amener également à céder : l'enfant perd tout intérêt véritable et finit par accepter un seul bonbon uniquement en vertu du réflexe consistant à saisir une possibilité immédiate plutôt qu'une possibilité future²⁹⁰.

De bonnes compétences d'autorégulation jouent un rôle essentiel dans la stabilisation de tout comportement adapté et on peut raisonnablement s'attendre à ce qu'elles puissent atténuer les effets d'une sensibilité au rejet élevée. Elles pourraient intervenir à plusieurs niveaux : dans la phase d'anticipation du rejet, il est possible de contrecarrer la tendance à interpréter tout indice ambigu comme un signe de rejet. On peut résister à la tentation de voir une discussion animée comme un signal que la relation va bientôt terminer : il peut s'agir d'un échange salutaire d'opinions ou plutôt d'un accident causé par le stress dû à une période ou une situation difficiles. Après une véritable manifestation de rejet, un bon contrôle de l'attention peut être utilisé pour ne pas se focaliser sur la rage éprouvée mais plutôt sur l'objectif plus important de préserver la relation à long terme. Plusieurs expériences ont montré que de bonnes compétences d'autorégulation sont associées à plus d'estime de soi, de meilleurs résultats à l'école et au travail, plus de succès dans les relations sociales et un moindre usage de drogues. La corrélation entre une mesure d'autorégulation (délai de gratification ou autre) à un moment donné et des mesures de ces variables dans des questionnaires remplis successivement (par les sujets eux-mêmes, les parents ou des enseignants) reste significative à distance de plusieurs années. Le délai de gratification garde un pouvoir de prédiction après 20 ans²⁹¹.

RS et personnalité borderline

La personnalité bipolaire est un trouble psychologique qui affecte 1% à 2% de la population générale et entre 10% et 20% de la population clinique. Le trouble est caractérisé par un

²⁹⁰ J. Metcalfe et W. Mischel, "A hot/cool-system analysis of delay of gratification: Dynamics of willpower," *Psychological Review* 106, no. 1 (1999).

²⁹¹ O. Ayduk et al., "Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity," *Journal of Personality and Social Psychology* 79, no. 5 (2000).

comportement impulsif et masochiste (usage de drogues, dépression, boulimie), des fluctuations extrêmes dans l'humeur et une perception instable de soi-même et des relations, ce qui produit une alternance d'efforts désespérés pour se faire accepter et de manifestations inappropriées d'hostilité²⁹². Il n'est pas difficile de remarquer des similitudes entre la sensibilité au rejet et la personnalité bipolaire, d'autant plus que des niveaux modérés (non-cliniques) du trouble sont suffisants pour causer des problèmes d'adaptation professionnelle et relationnelle²⁹³. En particulier : (1) la peur du rejet est l'un des facteurs diagnostiques de la personnalité borderline dans le DSM-IV²⁹⁴ ; (2) la majorité des sujets borderline ont un style d'attachement évitant²⁹⁵ ; (3) il est probable que la sensibilité au rejet et la personnalité borderline ont des origines communes dans des expériences d'abus et rejet dans l'environnement familial, car dans les deux cas des symptômes précurseurs dans l'enfance sont plus communs chez des enfants maltraités que chez des groupes de contrôle²⁹⁶. Ces résultats sont cohérents avec ce qu'on sait du lien entre sensibilité au rejet et autorégulation : la corrélation entre les symptômes des deux conditions existe uniquement chez des sujets qui ont un bas niveau de compétences d'autogestion, et on vient de voir que l'autorégulation est capable d'atténuer les effets de la sensibilité au rejet chez des sujet qui néanmoins manifestent des réactions physiologiques typiques du trouble. S'il n'est pas question de proposer une simple identification entre la personnalité borderline et le type RS, ce dernier peut représenter un aspect du trouble mieux défini et testable avec davantage de rigueur empirique.

RS et dépression

Des dynamiques psychologiques de sensibilité au rejet jouent un rôle important dans le déclenchement d'épisodes dépressifs. La dépression est étroitement liée à des expériences de perte dans des relations sociales, mais toute forme de perte ne produit pas des épisodes dépressifs. Les théories existantes sont plutôt vagues sur les dynamiques cognitives précises qui expliqueraient le

²⁹² Ö Ayduk et al., "Rejection sensitivity and executive control: Joint predictors of borderline personality features," *Journal of Research in Personality* 42, no. 1 (2008).

²⁹³ C. Bagge et al., "Borderline personality disorder features predict negative outcomes 2 years later," *Journal of Abnormal Psychology* 113, no. 2 (2004).

²⁹⁴ Le DSM-IV (*Diagnostic and Statistical Manual - Revision 4*) est un outil de classification qui représente le résultat actuel des efforts poursuivis depuis une trentaine d'années aux États-Unis pour définir de plus en plus précisément les troubles mentaux.

²⁹⁵ H. R. Agrawal et al., "Attachment Studies with Borderline Patients: A Review," *Harvard review of psychiatry* 12, no. 2 (2004).

²⁹⁶ F. A. Rogosch et D. Cicchetti, "Child maltreatment, attention networks, and potential precursors to borderline personality disorder," *Development and Psychopathology* 17, no. 04 (2005).

déclenchement. Dans la tradition d'études sur le style d'attachement, la prédisposition à la dépression est associée à un style évitant, qui favoriserait entre autres un manque de confort dans l'intimité, une anxiété par rapport à l'abandon et un manque de confiance envers l'autre²⁹⁷. Comme nous l'avons remarqué, la liste des dynamiques psychologiques associées aux styles d'attachement est ouverte et les limites sont floues. Dans leur ensemble, elles sont censées créer une prédisposition à la dépression, mais aucune hypothèse plus précise n'a été avancée. D'autres approches insistent sur le fait qu'un investissement exagéré dans les relations et une anxiété liée au besoin de se faire accepter sont les antécédents les plus significatifs d'un épisode dépressif²⁹⁸. On retrouve ici le même problème : une constellation floue de dynamiques psychologiques et l'impossibilité de tester des processus cognitifs précis. Notamment, il est impossible de formuler des prédictions sur exactement quel genre d'expériences de perte vont être à la base de la dépression : ces théories postulent des facteurs déclenchants généraux tels que le « stress relationnel » et n'opèrent pas de distinctions aussi basiques que celle entre la fin volontaire d'une relation ou un abandon.

Par contre, la sensibilité au rejet permet de prédire une réaction dépressive seulement dans le cas où l'individu (1) s'attend à être rejeté (doute de son acceptation) ; (2) attribue beaucoup d'importance à la relation²⁹⁹. Une forte anticipation du rejet ne devrait pas être suffisante pour déclencher un épisode dépressif majeur si l'individu n'attribue pas une valeur spéciale à la relation : par exemple, s'il se trouve qu'il aurait probablement décidé de mettre fin à la relation de toute façon. Réciproquement, une personne qui mise tout sur une relation mais qui en général s'attend à ne pas être rejeté ne réagira pas de façon dépressive. Pour résumer, la dépression serait le résultat d'un échec dans un objectif relationnel très important pour lequel l'individu a fait le maximum de ses efforts. Les données montrent que des femmes RS manifestent plus de symptômes dépressifs qu'un groupe de contrôle après une séparation voulue par leur compagnon, mais pas après une séparation volontaire ou consensuelle, ni face au stress ou à un échec à l'université. Cela confirme le pouvoir prédictif de la sensibilité au rejet³⁰⁰.

²⁹⁷ K. D. Rudolph, C. Hammen et D. Burge, "A Cognitive-Interpersonal Approach to Depressive Symptoms in Preadolescent Children," *Journal of Abnormal Child Psychology* 25, no. 1 (1997).

²⁹⁸ J. C. Coyne et V. E. Whiffen, "Issues in personality as diathesis for depression: the case of sociotropy-dependency and autonomy-self-criticism," *Psychological Bulletin* 118, no. 3 (1995).

²⁹⁹ O. Ayduk, G. Downey et M. Kim, "Rejection Sensitivity and Depressive Symptoms in Women," *Personality and Social Psychology Bulletin* 27, no. 7 (2001).

³⁰⁰ Ibid.

RS liée au statut social

La sensibilité au rejet joue probablement un rôle décisif dans la psychologie des minorités qui font l'objet d'une discrimination. Des phénomènes d'intériorisation et d'anticipation de stéréotypes raciaux sont connus depuis des décennies. Lorsqu'un groupe fait l'objet d'une discrimination, ses participants souvent intériorisent le stéréotype concerné (ils se comportent de façon à le rendre vrai) ou projettent et anticipent des épisodes de discrimination à partir d'interactions neutres ou ambiguës. Dans le premier cas, il s'agit du phénomène de la « menace du stéréotype » (*stereotype threat*) : après avoir lu un article qui explique que les filles n'ont pas de résultats aussi bons que les garçons en mathématique, des groupes de filles ont en effet des résultats moins bons qu'un groupe de filles de contrôle³⁰¹. Dans le deuxième cas, il s'agit de la « conscience de la stigmatisation » (*stigma consciousness*) : que la discrimination soit effectivement probable ou non, le peur de s'y exposer peut pousser des femmes à éviter toute situation où un épisode de discrimination lié au sexe est possible³⁰².

L'hypothèse d'une sensibilité au rejet liée au groupe est très plausible : encore une fois, la valeur ajoutée de l'approche socio-cognitive réside dans la précision et testabilité du modèle, ainsi que dans l'effort d'étudier le processus psychologique vraiment à l'échelle de l'individu. La recherche sur la menace du stéréotype et la conscience de la stigmatisation se limite à la comparaison des moyennes entre groupes, alors que l'ambition socio-cognitive est celle d'expliquer les différences individuelles à l'intérieur des groupes qui font l'objet d'une discrimination. La sensibilité au rejet liée au statut est censée être indépendante par rapport à la sensibilité au rejet personnelle : pour tester cette hypothèse chez des étudiants noirs d'une université américaine, Downey et ses collègues ont développé un questionnaire qui a la même structure que le questionnaire standard, mais dans lequel les situations pertinentes décrivent des interactions avec des membres du groupe « discriminateur ». Notamment, il s'agit surtout d'interactions avec des personnes qui ont un rôle social important mais ne sont pas proches de l'étudiant (professeur, autorité publique). Au lieu de tester la réaction à l'idée de « demander une grande faveur à un ami proche » (RS standard), on demande de prédire le niveau d'anxiété éprouvée et d'estimer la probabilité d'épisodes tels que « être arrêté pour un contrôle à cause de sa couleur de peau » (respectivement sur deux échelles de 1 à 6). Les résultats du test montrent que cette mesure de la sensibilité au rejet liée au statut est indépendante de la sensibilité au rejet standard, de l'estime de soi et du niveau d'identification avec

³⁰¹ C. M. Steele, "A Threat in the Air:: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist* 52, no. 6 (1997).

³⁰² E. C. Pinel, "Stigma consciousness: the psychological legacy of social stereotypes," *Journal of Personality and Social Psychology* 76, no. 1 (1999).

sa propre ethnie. L'effet est dissociable d'autres manifestations de peur du rejet : quelqu'un peut être sûr de lui, mais avoir également peur d'être rejeté à cause de son groupe ; il peut ressentir une identité ethnique très forte mais ne pas anticiper le rejet, ou au contraire anticiper une discrimination tout en étant à l'aise dans les relations les plus significatives (RS standard). Néanmoins, la mesure est fortement corrélée à des questionnaires indépendants qui estiment respectivement l'anticipation d'un rejet sur des bases raciales et des réactions extrêmement hostiles à de vrais épisodes de discrimination, c'est-à-dire les deux dynamiques fondamentales de la sensibilité au rejet (anticipation, réaction hostile)³⁰³.

Un groupe d'étudiants universitaires de première année a été testé pour mettre à l'épreuve le pouvoir de prédiction du questionnaire. Il s'agissait d'écrire un journal personnel pendant 21 jours et de répondre à des questionnaires après un an. Dans le journal et le questionnaire, on demandait aux étudiants d'enregistrer leurs sensations et leurs opinions sur plusieurs aspects de la vie universitaire : l'humeur, le sentiment d'appartenance à l'université, l'attitude envers les professeurs et les autres étudiants, les expériences personnelles. Les étudiants noirs qui avaient une RS élevée liée au statut étaient caractérisés par une sensation de bien-être moins importante, une perception de rejet, moins de confiance pour les autres étudiants et les professeurs et moins d'amis blancs après un an. Ces résultats sont spécifiques aux étudiants d'origine afro-américaine et ne se retrouvent pas chez les Asiatiques, par exemple³⁰⁴.

La sensibilité au rejet a été notre étude de cas. Citons rapidement d'autres exemples de typologies socio-cognitives, pour mieux montrer le potentiel de l'approche.

Narcissisme

Les narcissiques partagent les biais cognitifs consistant à être exagérément attentifs à des menaces (réelles ou hypothétiques) à l'estime de soi, et à réagir de façon à neutraliser ces menaces³⁰⁵. En ce qui concerne la perception, une étude récente a montré que – après un stimulus subliminal lié à l'échec – les narcissiques sont beaucoup plus rapides qu'un groupe de contrôle dans un test de

³⁰³ R. Mendoza-Denton et al., "Sensitivity to status-based rejection: Implications for African American students' college experience," *Journal of Personality and Social Psychology* 83, no. 4 (2002). Etudes 1 et 2.

³⁰⁴ Ibid. Etude 3.

³⁰⁵ Le terme « narcissisme » tel qu'il est utilisé ici est un terme technique défini dans le cadre de l'approche socio-cognitive, même si sa signification se rapproche de la signification courante et la récupe largement. Pour un passage en revue de cette littérature (qui ne s'occupe pas des cas cliniques), voir F. Rhodewalt et D. L. Sorrow, "Interpersonal self-regulation: Lessons from the study of narcissism," *Handbook of self and identity* (2003).

reconnaissance lexicale (juger si une suite de symboles est ou non un mot) pour les mots qui expriment une baisse de l'estime de soi³⁰⁶. L'effet était spécifique et ne s'étendait pas à des mots négatifs ou neutres, montrant qu'il y a un lien cognitif fiable entre l'échec et le manque d'estime de soi. Lorsque le narcissique rencontre quelqu'un de plus capable que lui dans une tâche quelconque, il cherche systématiquement à le rabaisser, en sorte que la comparaison soit moins menaçante³⁰⁷. Le narcissique profite aussi de toute occasion pour s'élever et pour afficher ses qualités, tout en faisant très attention à gérer les risques que cela comporte. Par exemple, dans une expérience des sujets narcissiques profitaient d'une interview pour donner la meilleure image possible de soi, sauf lorsque ils savaient qu'ils faisaient face à un expert (ce qui augmente la possibilité d'être démasqué)³⁰⁸. Les narcissiques en général choisissent des situations, des travaux et des fréquentations compatibles avec leur attitude cognitive. Ils affichent davantage d'intérêt pour un jeu s'il est interprété comme étant compétitif plutôt que comme un défi envers soi-même, ou simplement comme une façon de s'amuser³⁰⁹. Pour résumer, le type narcissique se caractérise, au niveau de perception sociale, par une attention exagérée aux menaces à l'estime de soi et, au niveau de réaction, par des buts de auto-promotion et un désir de manifestation de supériorité. Encore une fois, l'intérêt de cette typologie réside dans la précision et la testabilité des mécanismes postulés.

Personnalité dépendante

Un dernier exemple de typologie socio-cognitive est la personnalité dépendante. Un sujet de ce type cherche l'assistance et le soutien des autres de façon exagérée et même dans des contextes où il aurait les moyens d'être autonome. D'un point de vue de la perception des situations sociales, un sujet dépendant est hyper-sensible aux signes de disponibilité de ressources et d'attention. Par exemple, dans une expérience de reconnaissance lexicale, après un stimulus subliminal lié à la disponibilité de support, les sujets dépendants étaient plus rapides qu'un groupe de contrôle à reconnaître des mots liés à la dépendance. Le fait que la stimulation subliminale n'ait pas produit de

³⁰⁶ S. Horvath et C. C. Morf, "Implicitly activated worthlessness in narcissism" (Données non publiées: University of Bern, 2006).

³⁰⁷ C. C. Morf et F. Rhodewalt, "Narcissism and Self-Evaluation Maintenance: Explorations in Object Relations," *Personality and Social Psychology Bulletin* 19, no. 6 (1993).

³⁰⁸ C. C. Morf, M. Davidov et D. Ansara, "Strategic regulation of desired identity images in response to opportunities and threats: Effects of audience characteristics on narcissistic self-presentation," *Manuscript under review* (2006).

³⁰⁹ C. C. Morf, C. Weir et M. Davidov, "Narcissism and Intrinsic Motivation: The Role of Goal Congruence," *Journal of Experimental Social Psychology* 36, no. 4 (2000).

changements dans la reconnaissance d'autres catégories de mots (chez aucun groupe de participants) est une preuve de la spécificité de l'effet³¹⁰. Dans une autre étude, les sujets dépendants étaient disposés à attendre plus longtemps qu'un groupe de contrôle pour commenter les résultats positifs d'une épreuve avec un professeur qui était décrit comme étant disponible pour aider et soutenir l'étudiant dans le futur. Lorsqu'ils étaient stimulés de façon subliminale (comme dans l'expérience précédente), ils pouvaient attendre encore davantage, mais toute différence avec le groupe de contrôle disparaissait lorsque le professeur était décrit comme n'ayant pas beaucoup de pouvoir académique et étant très occupé. Les sujets dépendants adoptent une série de stratégies pour s'assurer le soutien qu'ils cherchent : ils peuvent être passifs et complaisants dans une relation personnelle ou assertifs et prêts à se mettre en avant dans une relation avec des personnes qui ont du pouvoir et de l'autorité. Ils formulent souvent des requêtes exagérées et menacent de rupture de la relation (et même de se suicider) au cas où ils n'obtiendraient pas satisfaction³¹¹.

Il y a une certaine ressemblance entre la personnalité dépendante et la sensibilité au rejet, mais le type RS cherche plutôt à minimiser les chances de rejet et non pas à exiger une quantité disproportionnée de soutien (en quantité et en qualité). Au contraire, le sujet RS tend à éviter toute requête, de peur qu'elle soit rejetée. Comme pour la sensibilité au rejet, la personnalité dépendante est à la base d'une prophétie auto-réalisatrice : la succession de requêtes exagérées finit par engendrer de la frustration, de l'exaspération et de l'hostilité chez les proches, qui souvent commencent à exploiter ou contrôler les sujets dépendants, mais qui en aucun cas ne fournissent du support authentique. Cet effet auto-réalisateur prive le sujet de toute expérience qui pourrait remettre en question son schéma du soi en tant que faible et dépendant.

La cohérence de la personnalité à un niveau supra-individuel : les dynamiques dans une relation et la dimension culturelle.

Si l'objectif principal des théoriciens socio-cognitifs est la modélisation de la personnalité à l'échelle individuelle, les moyens théoriques développés nous permettent aussi de conceptualiser la cohérence comportementale au niveau d'entités supra-individuelles : depuis une relation entre deux personnes jusqu'à des régularités dans le comportement des gens qui partagent la même culture. Le

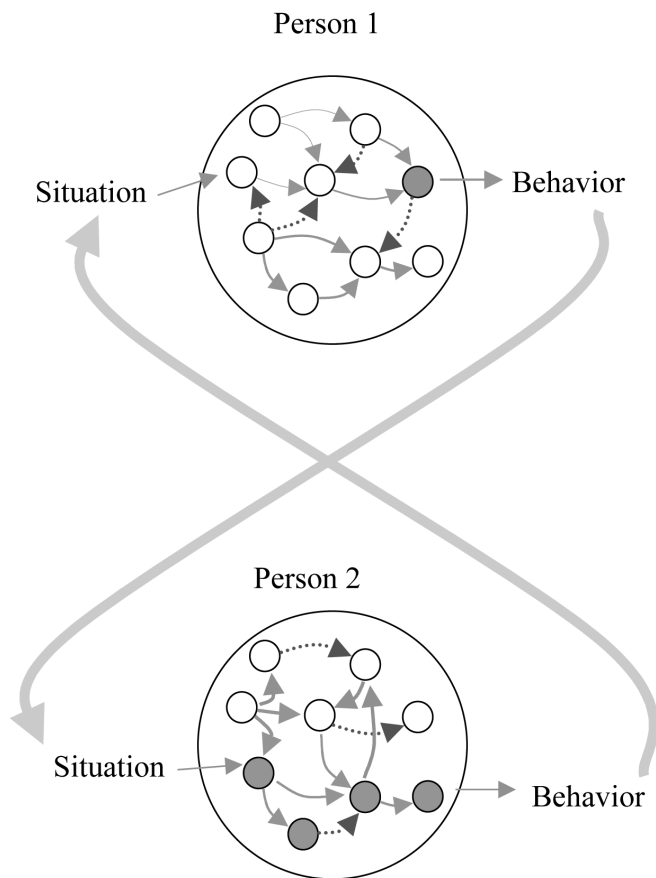
³¹⁰ R. F. Bornstein et al., "Contrasting Effects of Self-Schema Priming on Lexical Decisions and Interpersonal Stroop Task Performance: Evidence for a Cognitive/Interactionist Model of Interpersonal Dependency," *Journal of Personality* 73, no. 3 (2005).

³¹¹ R. F. Bornstein et al., "Activity, Passivity, Self-Denigration, and Self-Promotion: Toward an Interactionist Model of Interpersonal Dependency," *Journal of Personality* 64, no. 3 (1996).

cas d'une relation de couple ou d'une amitié profonde peut être traité par une simple extension du modèle connexionniste : nous avons vu que la cohérence dans les réactions émotionnelles et le comportement est représentée par l'ensemble des attracteurs dans lesquels le réseau de personnalité se stabilise dans chacune des situations sociales saillantes. Si l'attracteur est « large » l'attitude correspondante (par exemple : timidité) sera affichée dans un nombre très important de situations, si l'attracteur est « petit », il n'y aura qu'un nombre limité de situations en mesure de générer la réaction spécifique. Or, ici les attracteurs correspondent à un état du réseau de personnalité de l'individu, mais ne pourrait-on pas représenter l'interaction habituelle entre deux personnes par le biais d'un réseau formé par l'union (ou le branchement) des deux réseaux de personnalité ? L'hypothèse est beaucoup moins spéculative qu'elle ne le paraît et peut être prise au sens littéral : en fait, normalement, les inputs externes qui influencent le système de personnalité sont des situations sociales saillantes (être menacé, être soumis à des pressions sociales, faire l'objet d'un regard embarrassant, etc.) et, dans une interaction habituelle, la situation la plus saillante est déterminée souvent par l'action de l'autre. Le fait que mon meilleur ami soit fâché, particulièrement heureux ou déçu, me fasse une proposition stimulante ou se taise de façon embarrassante est évidemment saillant pour la gestion de la relation. Ainsi, dans une interaction, la situation est avant tout « l'autre ». Cela étant réciproque, les deux systèmes de personnalité peuvent être littéralement « branchés », l'action de chacun représentant une situation spécifique qui pousse le système de personnalité de l'autre vers un certain équilibre, et vice-versa. Intuitivement, on remarque souvent qu'une relation consolidée a presque « sa propre personnalité » : il y a des dynamiques comportementales stables à l'échelle de la relation. Cela ne se réduit pas au constat selon lequel chacun exerce une influence régulière sur l'état d'esprit de l'autre (je suis particulièrement inspiré par la présence de mon meilleur ami ; la mienne l'amène à se détendre). On peut assister à la synchronisation de deux réactions différentes face au même stimulus (dans une situation embarrassante, typiquement je commence à rire et mon ami s'inquiète et cherche une solution) ou à la production typique de séquences de réactions précises (lorsque je commence à m'énerver, il s'amuse ; je m'énerve encore plus ; il commence à se préoccuper ; je commence à me détendre, etc.). Ces régularités ne se produisent que dans le cadre de l'interaction : je ne peux pas m'empêcher de rire dans une situation embarrassante, mais uniquement en présence de mon meilleur ami ; c'est uniquement face à son attitude préoccupée que je commence à me calmer, alors que l'inquiétude d'autrui normalement ne me tranquillise pas.

Le modèle socio-cognitif est en mesure de prédire cet effet : dans une simulation de deux systèmes de personnalité branchés de la façon décrite (l'output d'un système est l'input principal de l'autre, et vice-versa) on retrouve des attracteurs du réseau global qui ne résultent pas de l'union des

attracteurs que les deux systèmes manifestent lorsqu'ils sont pris séparément (système 1 dans son attracteur a, système 2 dans son attracteur b)³¹².



Cela signifie que les qualités de précision, pouvoir prédictif et testabilité indépendante de l'approche socio-cognitive pourront être mises à contribution dans l'étude des dynamiques inter-individuelles. À une échelle plus large, on peut conceptualiser la dimension culturelle de la personnalité. Nous avons vu qu'un profil de personnalité (*personality signature*), avec la distribution des connexions et leur intensité exacte, est unique. Il n'y a pas deux personnes qui réagiront de façon très semblable dans *toutes* les situations. Néanmoins, il y a des types : deux personnes qui ont des niveaux élevés de sensibilité au rejet partagent des dynamiques cognitives très semblables et manifesteront des *patterns* de cohérence comportementale comparables. Il y aura toujours un facteur idéographique non-éliminable : par exemple, une situation précise dans laquelle une personne RS n'aura pas de difficulté à demander une faveur à un ami. Mais, en général, on peut prédire de façon fiable que les RS anticipent avec anxiété tout rejet possible. Par le même processus d'abstraction qui permet de définir des types, on peut essayer d'isoler les dynamiques cognitives que partagent les membres du même groupe ou culture.

³¹² Shoda, LeeTiernan et Mischel, "Personality as a Dynamical System: Emergence of Stability and Distinctiveness from Intra- and Interpersonal Interactions," 322.

On a déjà expliqué ce que le modèle des traits généraux prédit quant à la variation culturelle de la personnalité (voir chapitre sur la théorie psychologique des traits). L'objectif des défenseurs du *big five* est d'abord celui de retrouver les mêmes macro-facteurs chez des cultures différentes. Retrouver les *big five* dans une culture signifie prouver qu'il est possible de catégoriser les individus de façon stable et cohérente dans les cinq dimensions. En refaisant l'analyse factorielle à partir des résultats des questionnaires dans la langue locale, on devrait pouvoir vérifier que les gens ont un placement stable dans des dimensions telles que le névrotisme ou extraversion. Ces macro-traits doivent aussi être relativement indépendants entre eux (« orthogonaux » : pas de corrélation trop forte). Ensuite, il s'agit de prouver l'existence de différences entre les niveaux moyens des macro-traits chez des groupes différents, par exemple, les Français pourraient se révéler plus névrotiques que les Américains. C'est la version contemporaine la plus sophistiquée de l'idée classique des caractères nationaux (ou typiques des groupes de référence). Si on se limite aux tendances moyennes, l'objectif d'étudier les caractères culturels ou nationaux est parfaitement justifié et les moyens des théoriciens des traits adéquats³¹³. Mais les limitations sont exactement les mêmes qu'au niveau individuel, notamment l'impossibilité d'étudier la variation interne au groupe de référence. Au niveau individuel, si on trouve une association entre l'extraversion et un effet comportemental, on ne pourra pas savoir en réalité quels sous-groupes de la population de référence sont vraiment concernés. Au niveau du caractère national, on ne pourrait pas expliquer pourquoi certains individus manifestent la tendance typique beaucoup moins que d'autres (voire pas du tout). Sur ce point, l'approche socio-cognitive montre son avantage. La dimension culturelle de la personnalité est conceptualisée tout simplement comme une partie du réseau de personnalité qui serait partiellement isomorphe chez tous les membres du groupe.

On peut imaginer une série de sous-réseaux culturels finement intégrés dans la structure de chacune des personnalités individuelles. Cela signifie que, d'un point de vue cognitif, une culture serait caractérisée par un ensemble de manières de percevoir, traiter et réagir à des stimuli saillants. L'interaction entre les éléments culturels et individuels d'une personnalité permet d'expliquer la variation inter-culturelle dans la disposition à manifester le caractère typique du groupe. D'abord, il y a une dimension de spécificité situationnelle : les réactions typiques d'une culture sont activées dans certains contextes et pas dans d'autres. Un bon exemple est l'étude faite par Mendoza-Denton des réactions à la condamnation de O.J. Simpson dans la communauté afro-américaine aux Etats-

³¹³ Il reste le problème suivant : tous les macro-facteurs du big-five n'ont pas été retrouvés dans toutes les cultures, et plusieurs cultures semblent avoir des macro-traits locaux spécifiques, dits « émiques », voir A. T. Church, "Culture and Personality: Toward an Integrated Cultural Trait Psychology," *Journal of Personality* 68, no. 4 (2000).

Unis³¹⁴. Au premier abord, les réactions semblaient suivre une simple ligne de démarcation raciale, les blancs étant convaincus de la culpabilité de l'ancien champion et les noirs de son innocence. Mais la situation était plus complexe et intéressante d'un point de vue cognitif. Une analyse plus profonde a montré que certains aspects spécifiques du cas – notamment l'occultation des preuves par un détective de Los Angeles, pour favoriser l'incarcération – avaient activé le scénario d'un traitement injuste de la communauté afro-américaine de la part de la police (malheureusement, les précédents historiques sont nombreux). Selon l'analyse socio-cognitive de Mendoza-Denton, l'activation de ce scénario avait ensuite inhibé une évaluation objective des preuves à l'encontre de Simpson. Littéralement, les membres de la communauté afro-américaine n'arrivaient pas à se dire : « en effet, il y a beaucoup de preuves contre l'accusé ». Par contre, pour les blancs d'origine européenne la réalité des discriminations policières était beaucoup moins saillante (même s'ils pouvaient en être conscients à un niveau réflexif). Ils finissaient par se concentrer sur les preuves telles qu'elles se présentaient, ce qui a engendré la conviction que l'accusé est coupable. La remarque suivante est décisive : dans d'autres procès très médiatisés où l'accusé était noir, l'opinion publique ne s'était pas divisée selon des lignes raciales.

Cet exemple montre que les tendances générales et moyennes d'un groupe ne peuvent pas rendre compte de façon adéquate de la cognition culturelle. Les sous-réseaux culturels s'activent à partir de situations déclenchantes spécifiques : mobiliser les intérêts du groupe n'est pas suffisant. La dimension culturelle de la personnalité est caractérisée par la même complexité que la dimension individuelle. Dans les deux cas, il est impossible de prédire le comportement si l'on ne comprend pas qu'il s'agit de l'effet émergent d'un réseau d'unités affectives et cognitives.

³¹⁴ R. Mendoza-Denton et al., "Cognitive-Affective Processing System Analysis of Reactions to the OJ Simpson Criminal Trial Verdict," *Journal of Social Issues* 53, no. 3 (1997).

Introduction troisième partie

Après avoir établis les conditions de possibilité d'un dialogue entre théorie philosophique de la vertu et psychologie empirique contemporaine (première partie), et montré quelles sont les théories psychologiques les plus pertinentes pour cette démarche interdisciplinaire (deuxième partie), il s'agit maintenant d'évaluer les deux ambitions principales du modèle néo-aristotélicien des vertus : (1) la cohérence dans le comportement qui caractérise l'excellence dans un domaine pratique et (2) l'intégration et implication mutuelle entre plusieurs formes d'excellence (l'unité des vertus : le courage implique la générosité, la justice implique le courage, etc., si bien qu'en réalité les vertus sont intégrées et ne peuvent pas être développées séparément).

Les deux premiers chapitres seront consacrés à la contribution de deux traditions de recherche différentes au problème de la cohérence du comportement : la psychologie de la personnalité et la psychologie sociale situationniste. Ces deux courants s'intéressent au niveau de cohérence qui existe réellement chez les gens et cherchent à établir si l'état «naturel» du caractère humain est la fragmentation ou la cohérence. Pour cette raison, ils n'ont pas le dernier mot sur la faisabilité d'un changement de cet état naturel par le biais de l'entraînement ou d'une éducation spécifique. Il faudra, dans le troisième chapitre, analyser la contribution de la psychologie de l'excellence et de l'expertise, qui s'intéresse à la possibilité d'atteindre des performances exceptionnelles dans plusieurs domaines pratiques grâce à un entraînement particulièrement rigoureux. La psychologie a étudié l'excellence dans des activités pratiques, sportives et intellectuelles. Nous le verrons, ses conclusions sont remarquablement universelles. Malheureusement, les psychologues n'ont pas étudié directement l'excellence morale. Toutefois, appliquer la psychologie de l'expertise à la cognition morale pour expliquer la conduite des modèles moraux est, depuis toujours, la stratégie fondamentale de la théorie de la vertu. Même si un jour on découvrirait que la cognition morale fait exception par rapport au modèle général de l'expertise, la théorie de la vertu ne peut pas se permettre d'ignorer les données contemporaines sur la performance excellente, car elle est la première à avoir proposé ce parallèle et a depuis toujours adopté les modèles anciens de l'expertise. Or, la psychologie nous montre que l'excellence est également plutôt « fragmentée » : il est difficile d'être véritablement experts en plus d'un domaine (la double expertise est très rare) et, même dans ce domaine, le niveau de compétence excellent est beaucoup plus restreint qu'on ne le croirait. Par exemple, des médecins considérés comme excellents se distinguent en réalité uniquement dans le traitement d'une ou deux maladies. Les données empiriques mettent à mal aussi la thèse de l'unité et du renforcement mutuel entre vertus. L'excellence pratique est très difficile à

atteindre et l'entraînement requis produit des changements psychophysiques majeurs qui limitent le potentiel de plasticité de l'organisme. Atteindre l'excellence dans un domaine peut se révéler un obstacle pour la recherche de l'excellence dans un autre domaine.

Finalement, dans le dernier chapitre nous allons nous intéresser à une notion-clé dans la théorie de la vertu classique : la sagesse. Traditionnellement, cette vertu est censée jouer un rôle décisif dans l'intégration des autres vertus. Le jugement sage permet d'arbitrer dans les conflits entre des vertus (ex : attaquer l'ennemi de cette façon montre du courage, mais est-ce que c'est loyal ?) et garantit le développement d'un caractère harmonieux. Nous allons analyser les données empiriques sur le fonctionnement de la sagesse (une tradition de recherche peu connue mais qui existe depuis une trentaine d'années). Encore une fois, le phénomène de la fragmentation émerge là où on espérait trouver l'intégration : on peut être plus au moins sage dans un domaine par rapport à un autre. La sagesse n'est pas un savoir monolithique, que soit on possède soit on ne possède pas. Pour qu'elle puisse assurer le renforcement mutuel et l'intégration entre *toutes* les vertus, encore faudrait-il être sages dans tous les domaines de la vie. Etant donné que la sagesse est plurielle et difficile à atteindre, on retrouve ici le même problème de fragmentation qu'on cherchait à éviter au départ.

Il faudra se résoudre à abandonner le modèle néo-aristotélicien en faveur d'un modèle de vertu en tant que spécialisation morale. Les caractéristiques de ce modèle seront définies dans la conclusion de cette thèse. Résumons de façon plus détaillée le contenu de chaque chapitre.

Premier chapitre

Dans le débat contemporain en psychologie morale des vertus, on pose le problème de la fragmentation du caractère uniquement à partir des données de la psychologie sociale situationniste, qui montrerait l'influence surprenante de facteurs anodins de la situation sur l'attitude morale de l'agent. La psychologie de la personnalité, notamment la tradition des traits qui a abouti au modèle du *big five*, n'est pas prise en compte. Comme nous l'avons montré, la psychologie de la personnalité s'intéresse à des traits peu fiables et qui n'ont que peu de rapport avec la conception classique des vertus : la vertu est associée à la cohérence dans des habitudes complexes et « intelligentes », non pas à des automatismes de bas niveau ou à des procédures « scriptées », et encore moins aux simples tendances tempéramentales exprimées par le *big five*. En ce sens, les philosophes ont raison à croire que la psychologie de la personnalité contemporaine s'intéresse à autre chose, notamment à l'étude du simple tempérament, par opposition à la sophistication de l'excellence pratique dans un domaine moral. Mais cette façon de voir les choses est paradoxale et cache une profonde ignorance de l'histoire de la psychologie de la personnalité. En fait, si les

psychologues ont dû se rebattre sur l'étude de tendances tempéramentales qui sont faibles par rapport aux attentes de la philosophie, c'est que, au fil de presque un siècle de recherche, ils n'ont trouvé aucune façon de montrer l'existence d'une cohérence significative au niveau d'attitudes complexes, « intelligentes » et de haut niveau, alors que c'était là précisément leur but déclaré. Nous montrerons que les fondateurs de la psychologie de la personnalité avaient des théories du caractère très proches de la psychologie des vertus. Ils ont dû faire face très rapidement à la difficulté extrême de tester empiriquement ces théories, et cela a donné naissance à une tradition expérimentale qui – en première analyse – *semble* être beaucoup moins ambitieuse, car elle *semble* ne s'intéresser qu'à des régularités superficielles dans le comportement. En réalité, jusqu'à récemment (quand, avec le *big five*, le but est devenu *vraiment* moins ambitieux), les psychologues ont conservé leur intention d'étudier la cohérence d'attitudes complexes, et ils ont cherché à le faire par le biais d'un grand nombre de stratégies ingénieuses. Dans la foulée de cet effort, deux résultats philosophiquement intéressants sont à signaler. Premièrement, la psychologie de la personnalité a produit une clarification conceptuelle remarquable des différents types de cohérence que l'on peut rechercher dans le comportement, de leur intérêt, des liens qui subsistent entre ces concepts et de leur utilité. Ce travail, complètement ignoré pour l'instant, mérite d'être porté à l'attention des philosophes. Deuxièmement, l'ingéniosité des stratégies qui ont été déployées pour dénicher la cohérence transforme cette quête en une preuve négative : si on a cherché aussi bien sans trouver, c'est que la cohérence complexe recherchée n'existe pas. Ce n'est pas une preuve incontestable, mais un argument à prendre sérieusement en considération.

Deuxième chapitre – la critique du situationnisme radical

Les arguments les plus discutés dans le débat sur la fragmentation du caractère sont inspirés par la littérature situationniste. Les expériences « situationnistes » nous donnent effectivement des raisons ultérieures pour douter de la possibilité de la cohérence, mais, comme nous allons le montrer dans le deuxième chapitre, la force de ces arguments est largement surévaluée.

La stratégie situationniste comporte deux volets : démontrer une fragmentation radicale du caractère à partir des expériences, et expliquer l'illusion de la cohérence (pourquoi les gens ne se rendent pas compte de la fragmentation radicale ?). La preuve de la fragmentation radicale se base sur la généralisation d'un phénomène qu'on va appeler « piège situationnel » : supposons qu'un individu possède une attitude associée à un niveau important de cohérence, disons la générosité ; il sera toujours possible (et facile) d'imaginer une situation où sa générosité est censée se manifester mais où l'individu ne va pas agir de façon généreuse. En particulier, la différence entre une situation où

la générosité s'active et un piège situationnel tient souvent à un détail anodin de la situation, par exemple le fait d'être pressé, d'être en groupe plutôt que tout seul, d'être de bonne humeur, etc. En généralisant à partir d'une longue tradition expérimentale qui vise à poser des pièges situationnels aux participants, les situationnistes se sont persuadés qu'il est possible de réfuter la cohérence de n'importe quelle attitude par cette méthode. Même si le nombre de pièges situationnels qui ont été produits est en effet remarquable, une généralisation aussi extrême est problématique. L'idée selon laquelle des pièges situationnels nous encerclent systématiquement n'est pas étayée par une théorie psychologique intéressante et profonde : elle se base uniquement sur le constat qu'on n'a pas pour l'instant trouvé des limites à la possibilité de reproduire et multiplier de tels effets.

La généralisation est aussi problématique à cause de l'absence d'une bonne explication du fait que l'on ne se rende pas compte de la non-existence des traits larges, alors même que nous connaissons bien les incohérences de nos proches. L'explication fournie par les situationnistes mobilise le biais fondamental d'attribution : une tendance à expliquer le comportement d'autrui par l'attribution de tendances générales et cohérentes, même lorsque cela n'est pas du tout justifié. À partir d'une analyse de la littérature sur l'explication, nous montrerons que ce biais existe mais il se manifeste essentiellement dans l'interprétation du comportement des inconnus. Sans vouloir nier le rôle que la simplification et la projection de stéréotypes jouent dans la compréhension de nos proches, par ailleurs nous mobilisons aussi une connaissance détaillée de leur personnalité. Par exemple, nous savons très bien *a priori* que notre meilleur ami n'est pas timide, mais qu'il le devient avec les jolies femmes ou s'il doit appeler un inconnu au téléphone, que notre père est autoritaire mais qu'il n'oserait pas nous réprimander en présence de son vieil ami d'enfance, que notre grand-père est généreux mais qu'il ne faut pas lui demander d'argent le jour où il paye les impôts, que le professeur est normalement impartial mais qu'il vaut mieux ne pas faire devant lui des commentaires ironiques. Dans le chapitre sur les théories de la personnalité du sens commun, nous avons montré que toute une kyrielle de structure cognitives (scripts, schémas relationnels) sont spécifiquement consacrées à la compréhension des proches. L'illusion de la cohérence, alors que la réalité serait la fragmentation totale, n'existe pas. L'explication des situationnistes présuppose que le biais fondamental d'attribution soit à la base de toute la cognition sociale : cela n'est pas tenable. Les situationnistes exagèrent sans doute le niveau de fragmentation du caractère, car une fragmentation extrême ne pourrait pas échapper aux moyens sophistiqués que nous déployons pour comprendre les personnes avec qui nous partageons notre vie.

Cela dit, globalement, l'enseignement de la psychologie de la personnalité (sous-évalué) et les arguments situationnistes (surévalués) montrent que l'état naturel de la personnalité humaine est

effectivement caractérisé par un niveau très important (probablement pas radical ou extrême) de fragmentation.

Troisième chapitre – Vertu morale et psychologie de l'expertise

Dans le troisième chapitre, nous nous posons la question de savoir si l'état naturel de fragmentation du caractère peut être dépassé par un entraînement ou une éducation suffisamment efficaces. La possibilité d'atteindre un niveau d'excellence dans une activité pratique (telle que la guerre, le commerce, l'artisanat, la gestion des affaires publiques) grâce à un entraînement systématique a été depuis toujours la référence de la théorie de la vertu d'inspiration aristotélicienne. À l'état naturel, le caractère est fragmenté, certainement beaucoup plus que les anciens ne le croient. Mais cela ne pose pas forcément de problème majeur. En fait, les anciens croyaient déjà dans la nécessité d'un entraînement très long pour pouvoir atteindre la vertu. Cet entraînement était déjà spécifiquement conçu pour surmonter le manque d'harmonie et la fragmentation du caractère naturel. Que le problème à résoudre (la fragmentation) soit un peu plus grave que prévu ne change pas la donne. Par rapport à la stratégie de la théorie de la vertu, la psychologie contemporaine semble imposer un nouveau calibrage de certains paramètres, non pas l'abandon du modèle. Après tout, personne ne sera choqué d'apprendre que la vertu est un peu plus difficile à atteindre qu'on ne l'avait estimé.

Cependant, les théories psychologiques de l'excellence qui ont inspiré la théorie de la vertu ignoraient la distinction entre excellence et simple compétence ou maîtrise, établie par la psychologie contemporaine. Dans l'exercice de quelque activité, pratique ou intellectuelle que ce soit, il y a deux objectifs saillants qui marquent des tournants dans la courbe d'apprentissage : la compétence (ou maîtrise), et l'excellence. La compétence correspond intuitivement à un niveau acceptable et moyen de maîtrise d'une tâche : en d'autres termes, savoir vraiment faire quelque chose. Être un bon joueur d'échec, être diplômé en violon, savoir vraiment jouer au poker, avoir la capacité de diagnostic d'un médecin pratiquant moyen, etc. L'excellence est définie par l'appartenance au groupe des meilleurs pratiquants d'une activité : non nécessairement les dix champions les plus connus, mais il s'agit de très haut niveau.

Le niveau de compétence (ou maîtrise) peut être atteint après une période d'entraînement non nécessairement très longue ; en plus, des compétences différentes se renforcent mutuellement (savoir conduire et avoir une bonne connaissance de la géographie sont deux compétences qui s'intègrent chez un voyageur).

Jusqu'à il y a quelques décennies, la psychologie de l'excellence était considérée comme une extension de la psychologie de la maîtrise : une simple version extrême et plus poussée. C'est

également la position adoptée dans l'antiquité et qui a influencé la psychologie de la vertu : la différence entre un bon artisan et un artisan excellent est une question de degré et de quantité d'expérience, le modèle psychologique sous-jacent étant le même. Aujourd'hui, il paraît clair qu'il y a des différences psychologiques fondamentales entre les deux niveaux de performance. L'excellence demande un entraînement beaucoup plus spécialisé et long ; elle s'associe à une véritable transformation psychophysique profonde qui épuise le potentiel de plasticité de l'organisme. Il est très rare d'être excellent en deux domaines. Si les compétences se renforcent réciproquement, l'excellence dans un domaine pose des obstacles à l'excellence dans d'autres domaines, en termes de temps nécessaire (il n'y a pas de transfert facile, il faut passer par toutes les phases de l'apprentissage dans le nouveau domaine) et d'épuisement du potentiel de plasticité de l'organisme. Or, la vertu, parce qu'elle est si difficile à atteindre et parce que l'état naturel du caractère est la fragmentation, relève sans doute du domaine de l'excellence pratique. Le modèle psychologique de la compétence est parfaitement compatible avec les ambitions de la théorie de la vertu, car les compétences se renforcent réciproquement, tout comme les vertus dans le modèle néo-Aristotélicien. Mais le modèle psychologique de l'excellence impose plutôt un cadre de spécialisation morale.

Dans sa tentative de réponse aux critiques de la psychologie de la personnalité et de la psychologie sociale, la théorie de la vertu a un allié inespéré : la psychologie de l'automatisation. On a toujours su que les habitudes et les automatismes sont des processus cognitifs très peu coûteux en termes d'énergie, d'attention et de concentration, ils sont aussi extrêmement efficaces et fiables. On a tout avantage à automatiser un processus, dans toute la mesure du possible. Il vaut mieux cuisiner, conduire, et même planifier son emploi du temps en faisant confiance autant que possible à de bons automatismes. Or, jusqu'à il y a quelques décennies, les psychologues croyaient que la stratégie consistant à « tout automatiser » ne pouvait pas nous amener très loin, à cause d'une forte limitation dans la complexité cognitive des processus passibles d'automatisation complète. En d'autres termes, les habitudes et les automatismes ne seraient que de simples réflexes de bas niveau : des processus cognitifs plus complexes demanderaient l'intervention de la réflexion consciente. Mais la psychologie de l'automatisation contemporaine ne cesse de montrer la complexité des processus automatisables : tout ce qui devait relever de la réflexion consciente (ex. contrôle méta-cognitif de plusieurs processus automatiques, poursuite de buts sophistiqués) peut être pris en charge par « le pilote automatique ».

Face à ces développements, la théorie de la vertu a tout intérêt à mettre en avant une conception totalement automatisée de la vertu, selon laquelle le rôle de la réflexion chez l'individu vertueux est

minimal. L'intérêt stratégique de cette conception « intuitionniste » de la vertu est évident : (1) la vertu est difficile à apprendre ; (2) tout ce qui relève entièrement d'une perception intuitive ou d'automatismes devient facile, libérant des ressources cognitives pour le reste, donc, (3) la vertu a intérêt à être aussi intuitive que possible. La stratégie des défenseurs d'une conception fortement intuitionniste de la vertu est ingénieuse et peut compter sur le soutien inattendu de plusieurs résultats de la recherche contemporaine sur la cognition automatique. Cependant, la psychologie de l'expertise et de l'excellence montre que la super-puissance de la cognition automatique a ses limites. Il est vrai qu'il faut automatiser « autant que possible », mais la complexité et la charge cognitive associées à une performance excellente sont simplement trop importantes pour que la cognition automatique puisse prendre tout en charge. Considérons l'exemple d'un champion de tennis qui veut garder son niveau : la masse de nouveautés techniques et stratégiques introduites par les nouveaux arrivés dans le circuit international est telle que, à chaque moment donné, le champion ne pourra pas avoir tout intériorisé. Il devra aussi réfléchir et délibérer. Les données prouvent que la réflexion joue un rôle semblable dans tout domaine d'excellence.

L'excellence est également caractérisée par un niveau de flexibilité et d'adaptation qui n'est pas associée à la simple maîtrise. On retrouve chez l'expert excellent un contrôle fin et un calibrage exquis de ses propres automatismes. Cela génère des réactions créatives et adaptées à des situations nouvelles.

Chapitre 4 – l'intégration de la vertu et le rôle de la sagesse

Mettre en avant le rôle intégrateur joué par la notion de sagesse est exactement la stratégie de la théorie de la vertu pour contrer les objections qu'on vient d'exposer. La sagesse est une source d'intégration de l'excellence pratique dans des domaines différents: le sage sait comment résoudre les conflits entre le courage, la justice et la modestie. Peut-être, dans les domaines d'excellence pratique qui sont moralement neutres, on ne dispose pas des avantages offerts par la sagesse. La sagesse ne peut pas aider un physicien excellent à être aussi un écrivain excellent, ni un médecin spécialisé dans la chirurgie nasale à être un médecin généraliste excellent. Voilà donc la réponse de la théorie de la vertu: dans le domaine moral, la sagesse rend possible une intégration qui reste très rare ailleurs. Le quatrième chapitre est consacré à tester le bien-fondé empirique de cette stratégie de réponse. La conclusion sera négative: à l'aune de nos meilleures connaissances, la sagesse ne semble pas en mesure de résoudre le problème de la « fragmentation de l'excellence », qui est aussi grave dans le domaine moral qu'ailleurs. On peut être plus au moins sage dans un domaine par rapport à un autre. La sagesse n'est pas un savoir monolithique, que soit on possède soit on ne

possède pas. Pour qu'elle puisse assurer le renforcement mutuel et l'intégration entre *toutes* les vertus, encore faudrait-il être sages dans tous les domaines de la vie. Etant donné que la sagesse est plurielle et difficile à atteindre, on retrouve ici le même problème de fragmentation qu'on cherchait à éviter au départ. Cela ne laisse qu'une possibilité: un modèle de la vertu en tant que spécialisation morale.

A la recherche de la cohérence dans le comportement : la leçon de la psychologie de la personnalité pour la théorie des vertus

Si on appelle littérature « situationniste » au sens large l'ensemble des données empiriques qui ont été mobilisées pour critiquer les ambitions de la théorie de la vertu, dans ce chapitre nous allons analyser la contribution de la psychologie de la personnalité à la conceptualisation et à une exploration empiriquement élaborées du problème de la cohérence du comportement (un débat qui a presque un siècle). Le prochain chapitre sera consacré à la littérature situationniste « proprement dite », c'est-à-dire l'étude systématique de l'influence des situations développée par la psychologie sociale (un débat qui a une cinquantaine d'années).

I. L'étude de la cohérence inter-situationnelle (*cross-situational consistency*) et sa pertinence pour la théorie de la vertu : prémisses méthodologiques

Classification des formes différentes de cohérence

Toutes les critiques radicales de la théorie de la vertu basées sur la psychologie contemporaine (et pas seulement celle que nous allons développer dans cette section) mettent en cause une propriété bien précise censée caractériser le comportement humain : la cohérence inter-situationnelle. Par des stratégies argumentatives différentes, ces critiques convergent vers une mise en question de la cohérence qu'on peut afficher dans la vie de tous les jours. L'intuition de départ qui a motivé la recherche est que les gens que nous connaissons ne nous surprennent pas constamment, on a l'impression d'avoir suffisamment compris comment ils fonctionnent et on a à leur endroit toute une série d'attentes. « Il n'aura jamais le courage d'aller parler à une fille dans un bar », « il ne dira jamais non à une invitation pour une fin de semaine à la mer », « il va sans doute travailler beaucoup pour ses examens finaux ». On pourrait prolonger indéfiniment la liste pour chacun de nos proches et de nos connaissances : dès qu'on les connaît, on a l'impression de voir un bon niveau de cohérence dans leur comportement. Lorsqu'on croit être justifié dans l'attribution à un proche d'un trait comme la générosité ou le courage, on croit être sûr du niveau de cohérence dans son comportement qui justifie cette attribution.

Comme cela deviendra clair, le situationnisme se joue entièrement sur ce sujet : si l'on avait des raisons solides indépendantes d'affirmer que les gens sont cohérents, les arguments situationnistes

seraient neutralisés d'avance ; si, au contraire, la plausibilité de la cohérence était déjà compromise, ces arguments ne seraient pas nécessaires, car l'objectif aurait été déjà atteint. Il est donc très important de dépasser une compréhension intuitive et d'être le plus clair possible sur la notion exacte de cohérence qui fait problème.

A en juger par l'histoire de la question, cela ne va pas être facile. La formulation intuitive du problème paraît transparente : « Parler de traits de caractère et de vertus, de courage, d'honnêteté, d'ambition, de générosité ou de sociabilité (etc.) c'est parler à tout le moins d'une forme de cohérence dans le comportement. Parmi beaucoup d'autres choses plus complexes, lorsqu'on considère que quelqu'un est courageux on veut *au moins* dire que son courage va se manifester dans plusieurs contextes. Il va agir *comme le courage le requiert* dans un type de situation a et dans une type de situation b, c etc. » Une version plus articulée de ce qu'on veut dire intuitivement par cela doit faire nécessairement partie d'une définition minimale de « trait de caractère », mais comment va-t-on rendre plus précise et opérationnelle notre intuition pour pouvoir vérifier si la réalité des faits s'y conforme ? Il se trouve qu'après plus de 80 ans de recherche empirique visant directement ce problème et de débat en psychologie de la personnalité, une solution n'a toujours pas fait de consensus complet. L'immédiateté de l'idée selon laquelle quelqu'un de courageux doit intervenir aussi bien pour sauver un enfant qu'une vieille dame (autrement, on peut commencer à douter de son courage) cache un gouffre de problèmes méthodologiques, conceptuels et définitionnels. Le lien conceptuel de fond tient : pas de traits sans une forme de cohérence. Mais le genre de cohérence appropriée dépend de la définition de trait qu'on adopte, et toutes les données empiriques qui ont été recueillies ne sont pas également pertinentes selon la notion spécifique de cohérence qu'on va retenir. Pour pouvoir poser le problème de la cohérence du comportement de façon intelligible et avancer vers une solution il faut accepter un principe méthodologique :

Il faut choisir à quel type de cohérence du comportement on s'intéresse : on n'est pas confronté à une notion unique difficile à analyser, mais à plusieurs notions de cohérence irréductibles, qui *n'offrent pas les mêmes avantages et ne posent pas les mêmes problèmes* pour leur étude, et qui *ne sont pas forcément compatibles*. La question : « sommes-nous cohérents ? » n'a pas de sens. Il faut choisir à quel type de cohérence on s'intéresse.

Que signifie le fait que l'attitude d'une personne est courageuse de façon cohérente ? A Quelles conséquences opérationnelles et vérifiables peut-on s'attendre de ce fait ? D'abord, la structure logique de base du problème. Agir de façon cohérente par rapport à un trait ou à une attitude qu'on est censé avoir est faire ce que l'attitude requiert dans plusieurs circonstances pertinentes. Nous

pouvons introduire les termes techniques de *comportement requis* et de *situations fonctionnellement équivalentes*. Un courageux qui sauve un enfant et le jour suivant une vieille dame est cohérent parce qu'il a affiché le comportement requis (sauver) dans deux situations fonctionnellement équivalentes (par rapport au courage, les deux situations sont équivalentes en ce qui demandent la même chose, à savoir un acte de sauvetage). Si formellement la cohérence est donc une question de comportement requis dans des situations fonctionnellement équivalentes, toute détermination substantielle dépend de la nature du trait concerné. Dans la définition du trait, on a la plus totale latitude (mais rien n'assure que le trait ainsi défini existe et soit intéressant). Les situations pertinentes peuvent coïncider avec toutes les situations possibles, par exemple si on définit « nervosité » comme « afficher constamment une attitude nerveuse », il s'ensuivra qu'être cohérent revient à afficher sa nervosité tout le temps. Par contre, il y a des traits qui peuvent ne s'activer que très rarement : pour un vindicatif, la cohérence se mesure sur des épisodes épars.

Il y a quatre paramètres fondamentaux pour classer les genres possibles et non-équivalents de cohérence qui peuvent être mobilisés par le sens commun, les philosophes ou les psychologues selon les besoins explicatifs et descriptifs du moment. Les permutations de ces facteurs définissent autant de notions de trait profondément différentes :

1. détermination subjective ou objective des situations fonctionnellement équivalentes : un trait peut être défini par des comportements requis dans des situations spécifiées de façon objective, indépendamment de variations possibles dans l'interprétation personnelle de la situation, par exemple : « une vieille dame qui tombe dans l'eau » et « un enfant qui tombe dans l'eau » sont deux situations objectives fonctionnellement équivalentes pour le courage. Autrement, un trait peut se baser sur l'équivalence fonctionnelle entre des situations déterminées de façon subjective : on peut être irascible lorsqu'on se sent menacé, avec une perception très subjective de la menace. Un trait est objectif si tout le monde perçoit la situation de la même manière et réagit de façon appropriée, subjectif si on réagit de façon appropriée relativement à la façon dont on a perçu la situation. Attention à ne pas faire les deux confusions suivantes : **A.** la notion de trait objectif est souvent considérée comme incohérente, car déterminer « à la troisième personne » les situations pertinentes pour un trait reviendrait à analyser un trait comme la disposition à réagir d'une façon appropriée à partir de comment quelqu'un d'autre (l'expérimentateur, le philosophe qui analyse le trait, etc.) perçoit la situation, ce qui est bien sûr une absurdité psychologique. « Trait » est un terme psychologique et chaque personne ne peut que réagir à sa propre perception de la situation. La critique dévient vite une défense d'une conception subjective des traits : il serait absurde d'étudier comme les gens réagissent à des

situations catégorisées comme dangereuses par Aristote, il faut bien sûr étudier comment les gens réagissent à des situations qu'eux-mêmes considèrent subjectivement comme dangereuses. Cette critique ne naît que d'une confusion conceptuelle : bien sûr dans la psychologie réelle de chacun il n'y a réaction qu'à la manière dont lui-même perçoit la situation. Mais tout ce qu'on veut dire en considérant un trait x comme objectif (disons le courage), c'est qu'il y a une façon de percevoir les situations pertinentes (ex : un enfant est en train de noyer) telle que, si on ne la respecte pas (on voit les choses différemment), on ne possède pas le trait x (dans cette définition). Quelqu'un qui n'a pas peur, mais ne voit aucune raison d'intervenir face à un enfant en train de se noyer n'est pas courageux au sens objectif. La notion de trait objectif est parfaitement cohérente. Tout théoricien a le droit de définir un trait et défendre normativement son importance. Il n'y a aucun abus injustifié d'autorité : un trait objectif est défini uniquement par le fait que tout le monde catégorise les situations pertinentes de la même façon, on pourrait même découvrir des traits objectifs dont on n'avait pas soupçonné l'existence (que les théoriciens les apprécient ou pas). **B.** La notion de trait subjectif n'est pas empiriquement vide, ce n'est pas la tautologie selon laquelle tout le monde est courageux à sa façon, où n'importe quelle attitude serait appelée « mon courage à moi ». Il y a une variation individuelle dans la perception de quelle situation est dangereuse, mais il reste objectif et vérifiable que *si* la situation est perçue comme dangereuse, alors certaines réactions spécifiques seront déclenchées. En principe, le fait qu'une situation soit perçue comme dangereuse est indépendamment testable (transpiration, analyse de la salive, etc.) et donc il n'y a pas de circularité.

2. nature comportementale ou normative du comportement requis : le comportement requis peut être spécifié directement comme un type déterminé d'action (par exemple, quelqu'un de souriant dans les situations fonctionnellement équivalentes va tout simplement sourire, le trait « souriant » concerne par définition un seul type d'action) ou des types d'actions complètement différents peuvent être normativement requis par le trait (quelqu'un de sociable peut être amené à rire dans un contexte et à entamer une conversation dans un autre, il ne s'agit pas de rire tout le temps).
3. cohérence comparative/ cohérence absolue : le comportement requis peut être conceptualisé par rapport à une moyenne du groupe de référence (être agressif signifie se comporter de façon plus agressive que la moyenne dans la même situation ou en général) ou dans une échelle absolue (on juge que faire certaines choses – par exemple insulter – signifie être agressif, que les autres le fassent ou non)
4. cohérence homogène ou cohérence non-homogène : il s'agit de savoir si toutes les situations fonctionnellement équivalentes ont le même poids pour évaluer la cohérence globale (à

l'exception de contextes où il ne faut évidemment pas sourire, tout sourire nié compte de la même manière pour comprendre le niveau auquel on est souriant) ou si certaines situations sont plus importantes et centrales que d'autres (il est plus grave de ne pas aider un blessé à côté de nous que de ne pas faire l'aumône à l'énième personne qui nous la demande au fil du même jour).

Ce genre de classification ne concerne pas la question du statut de ces traits, s'ils existent réellement, s'ils sont explicatifs ou s'ils se réduisent à la reconnaissance de schémas dans le comportement. Quelle que soit la réponse sur ce sujet, ces traits sont légitimement invoqués pour des fins différentes et leurs propriétés ne sont pas équivalentes. On a le droit de se demander si une personne est souriante, voulant savoir par là seulement si elle sourit beaucoup en moyenne par rapport aux autres, il s'agira donc d'un trait objectif, comportemental, comparatif et homogène. Dans un autre contexte, par exemple la comparaison de l'attitude des gens de pays différents, on peut vouloir conceptualiser « souriant » par rapport à une échelle absolue de référence. Si on s'intéresse à l'irascibilité, on pense plutôt à une personne qui va réagir de façon plus violente et émotionnellement chargée que la normale lorsqu'elle se sent menacée ou mise en cause d'une façon nette : ce sera donc un trait subjectif, comportemental (quoique on puisse imaginer que dans certains contextes problématiques quelqu'un d'irascible va se taire et contrôler toute la colère, il s'agira donc d'une conceptualisation non comportementale du trait de l'irascibilité), comparatif et non-homogène (car une absence de réaction face à une mise en question légère, à peine ébauchée, ne va pas compter contre la cohérence de son trait au même niveau qu'un manque de réaction face à une menace nette). Cela n'a pas de sens de parler de la cohérence du comportement (sous-entendant qu'il y a un sens de cohérence objectivement plus important que les autres), car *ces types de cohérence différents sont utiles et intéressants pour des fins différentes.*

Une preuve expérimentale d'incohérence se base sur un paradigme qui a une structure formelle simple, il s'agit de décrire la mise en œuvre du comportement requis dans une situation fonctionnellement équivalente *a* et l'absence du comportement dans une autre situation fonctionnellement équivalente *b* (bien sûr, le diagnostic d'incohérence sera d'autant plus étayé que le phénomène est facilement répliquable et généralisable pour une personne donnée et sur un nombre plus important de situations). La réalisation pratique et l'interprétation ne posent pas les mêmes problèmes suivant la définition du trait concerné. *Ces types de cohérence ne sont pas équivalents en ce qui concerne la difficulté de leur étude.* En particulier, le caractère subjectif, normatif, absolu et non homogène ajoutent autant de couches de complexité méthodologique absentes dans l'étude d'un trait objectif, comportemental, comparatif et homogène. Si on part du cas le plus simple, le fait

d'être souriant défini comme une tendance générale à sourire, il s'agit juste de prouver que quelqu'un sourit dans une circonstance et pas dans une autre. On peut vouloir passer à une conception plus sophistiquée de l'habitude de sourire : par exemple, dans beaucoup de circonstances un sourire pourrait se révéler inopportun, il faudra donc les exclure des situations expérimentales, car ne pas sourire dans ces cas ne compte pas contre la cohérence de l'attitude. La couche de difficulté ajoutée consiste à être sûr de ne pas avoir inclus ces situations problématiques dans le paradigme.

Les complications ne font que commencer : si on s'intéresse à un trait *subjectif* tel que l'irascibilité, il faut contrôler que les gens voient subjectivement de la même façon les deux situations expérimentales, une provocation décrite en des termes objectifs à la troisième personne ne sera pas suffisante. Si le trait, comme il est vraisemblable pour l'irascibilité, n'est pas *homogène*, il faudra contrôler que les situations soient « centrales » et très importantes dans la structure du trait. Une interaction entre les deux problèmes est aussi inévitable : par rapport à une détermination objective à la troisième personne, des sujets différents peuvent avoir une perception non équivalente aussi bien de l'ensemble de situations qui sont pertinentes en général pour le trait que de l'ensemble de situations centrales et décisives pour sa manifestation. Un individu pourrait ne voir rien de mal dans une légère allusion ironique qui aurait au moins attiré l'attention d'un autre, alors que ce dernier va exploser face à un commentaire direct que le premier aurait pu tolérer en silence sans trop de difficulté. Enfin, déterminer la cohérence d'un trait *absolu* pose, comme nous le verrons, des problèmes méthodologiques d'interprétation des données, car les psychologues dans la plupart des expériences se sont limités à calculer des corrélations qui sont pertinentes uniquement pour la cohérence du positionnement relatif dans le groupe (plus d'éclaircissements sur ce point ci-dessous).

Quel genre de cohérence nous intéresse en ce qui concerne la vertu ? Le courant le plus représentatif et ambitieux dans l'histoire de la théorie de la vertu – et c'est bien ces ambitions qu'il faut prendre au sérieux tant qu'on n'aura prouvé qu'elles ne sont pas réalisables – conçoit les vertus comme un complexe et « large » (portant sur un vaste domaine d'action dans la vie) ensemble d'habitudes « intelligentes ». Des affirmations de ce genre sont très représentatives de l'approche aristotélicienne :

« Plus on développe une vertu, moins de simples habitudes deviennent importantes, et on développe davantage une capacité flexible et complexe de raisonner sur des genres de

*situations nouvelles et innovantes auxquelles on pourrait être confronté. Donc, plus on est vertueux, plus le caractère est complexe et dynamique*³¹⁵. »

Il ne s'agit pas d'automatismes ou de réflexes de bas niveau, mais de réactions de haut niveau : intelligentes, complexes et alimentées par une subtile sensibilité au contexte qui discrimine et évalue plusieurs alternatives d'actions. En abordant la littérature psychologique, il faut donc s'intéresser et rechercher une cohérence de processus psychologique de *haut niveau* sur un champ d'action *vaste*. Il existe bien sûr un courant minoritaire, psychologiquement neutre et quiétiste dans la tradition de réflexions sur la vertu³¹⁶. Par exemple, dans un cadre conséquentialiste Julia Driver définit la vertu comme n'importe quelle habitude qui a des conséquences généralement positives :

*« Ma théorie soutient simplement qu'un trait qui a de bonnes conséquences (good-producing) – acquis de n'importe quelle façon – est une vertu morale. Il pourrait être vrai de façon contingente que ces traits sont en général ceux qui exigent du travail dur pour être acquis et des efforts pour être maintenus*³¹⁷. »

*« Ainsi, la définition correcte de vertu morale sera celle qui a été développée ci-dessus, qui soutient simplement qu'une vertu morale est un trait de caractère qui produit du bien. Soutenir que la vertu morale est un trait de caractère est reconnaître que la vertu morale est quelque chose que ne peuvent posséder que des êtres qui ont un esprit ou un caractère. Ainsi, toute bonne attitude dispositionnelle n'est pas une vertu morale*³¹⁸. »

Ces remarques sont censées exclure a priori du domaine de la vertu uniquement des dispositions qui n'ont rien avoir avec les attitudes d'un sujet capable d'action intentionnelle et volontaire, par exemple des réflexes automatiques. N'importe quelle autre habitude qui relève de l'activité cognitive du sujet peut être vertueuse tant qu'elle a de bonnes conséquences. Une approche de ce genre est psychologiquement neutre et n'est pas directement concernée par notre analyse, qui s'intéresse à l'avantage comparatif pour une théorie de la vertu qui s'expose et mise sur une psychologie morale empiriquement adéquate. Il y aura forcément des habitudes qui ont des meilleurs conséquences que d'autres : la notion de vertu mobilisée par Driver n'est pas menacée par des découvertes empiriques. Le choix de nous centrer sur la possibilité de la cohérence associée à

³¹⁵ Julia Annas, "Comments on John Doris's 'Lack of Character'," *Philosophy and Phenomenological Research* 71, no. 3 (2005): 637.

³¹⁶ Merritt fait remonter à la théorie de la vertu de Hume cette tradition qui abandonne le perfectionnisme aristotélicien et considère la vertu davantage comme ce qui permet une vie morale décente plutôt que d'atteindre l'excellence, voir Merritt, "Virtue Ethics and Situationist Personality Psychology."

³¹⁷ Driver, *Uneasy virtue* 108.

³¹⁸ Ibid. 107.

une notion ambitieuse de vertu se justifie donc par (a) le fait que c'est l'option majoritaire et plus ancienne (b) la valeur des ambitions perfectionnistes qui ont animé un tel projet et qui ne doivent pas être écartées sans raison (c) la nature de notre enquête qui s'intéresse à des aspects de la théorie de la vertu où une base empirique solide peut faire la différence.

II. A la recherche de la cohérence inter-situationnelle : 80 ans de psychologie de la personnalité et leur pertinence pour la théorie de la vertu.

Toutes les stratégies qui ont été développées pour conceptualiser correctement et dénicher expérimentalement le genre de cohérence associée à des traits « larges » seront analysées. Contrairement à la perception de cette littérature répandue parmi les philosophes, nous allons montrer que les psychologues avaient bien l'intention d'étudier des phénomènes aussi complexes que ceux postulés par les théoriciens de la vertu, et que l'échec dans la recherche expérimentale de la cohérence de processus de haut niveau doit bien concerner les philosophes³¹⁹. Sans que l'intention d'étudier la vertu ait été jamais formulée directement, les stratégies déployées auraient dû aboutir si la vertu est répandue au-delà d'une minorité exiguë de la population.

L'enquête sur l'éducation du caractère et d'autres études classiques

Le débat sur la cohérence du comportement humain a été stimulé par une série d'études conduites dans la deuxième moitié des années 20 aux Etats-Unis. Motivés par la volonté de tester les présupposés psychologiques implicites dans le mouvement pour l'éducation du caractère, très développé et influent dans le système scolaire américain à cette époque-là, une équipe de psychologues (officiellement dirigés par les professeurs de Yale Hugh Hartshorne et Mark May mais en réalité guidés par le célèbre psychologue américain Edward Thorndike) ont administré plus

³¹⁹ Etant donné la complexité des théories des pères fondateurs de la psychologie de la personnalité (Allport, Cattell), la thèse selon laquelle les psychologues n'ont essayé de tester expérimentalement que des conceptions trop simplifiées des traits est sans doute fautive. Il est certainement difficile de tester ces théories complexes et – bien sûr – plusieurs expériences ont cherché à d'établir des résultats plus simples. Les philosophes – à partir des paradigmes expérimentaux isolés qui ont attiré leur attention – ont cru comprendre que les efforts des psychologues se concentraient sur des aspects trop spécifiques. Il faut prendre en considération une littérature expérimentale dans sa globalité : cumulativement, au fil des décennies la psychologie de la personnalité a testé des hypothèses complexes et raffinées sur le caractère.

de 170.000 tests dans plus de 8.000 écoles publiques et 2.500 écoles privées. L'objectif de ces études dont l'ampleur reste inégalé était de mesurer aussi bien les attitudes et les idéaux des enfants par rapport à la notion de caractère que le comportement effectif. Ce deuxième aspect, qui était d'ailleurs central dans l'intention des expérimentateurs, va nous intéresser.

Les résultats de ces efforts ont été publiés en trois volumes sur le titre de *Etudes sur la nature du caractère (Studies in the Nature of Character)*³²⁰. Le genre de paradigme typique de ces études va marquer l'histoire du débat : il s'agit de calculer la corrélation entre la conduite d'un groupe dans un contexte *a* et la conduite du même groupe dans un contexte différent *b* (ou la corrélation entre plusieurs contextes *a, b, c, d*). Si on fait exception de la possibilité idéale de suivre avec précision chaque personne d'un groupe nombreux pendant une très longue période de temps (ce qui est trop coûteux et n'a presque jamais été réalisé), cette méthode est la meilleure pour étudier la cohérence inter-situationnelle³²¹. Car, si le choix des situations est bon, l'attente d'une cohérence entre les deux relève directement de la définition d'un trait. Le deuxième meilleur paradigme est la corrélation entre le niveau d'un trait mesuré par des moyens psychométriques (questionnaires) et des comportements spécifiques en laboratoire ou dans un milieu naturel (le contexte et ce que l'agent fait doit être bien documenté). Par exemple, est-ce que les personnes très généreuses (d'après un questionnaire) s'arrêtent pour aider une vieille dame beaucoup plus que les autres ? Au choix d'une situation vraiment représentative (si, on regardant de près le contexte on découvre que la situation était ambiguë et que la vieille dame n'avait pas envie d'être approchée du tout, la force de l'expérience disparaît complètement), ici s'ajoute l'incertitude sur l'intérêt véritable du trait mesuré par le questionnaire, car une preuve d'incohérence est strictement relative à la définition du trait concerné, et aux chances que la méthode psychométrique soit valide. Néanmoins, si ces problèmes sont résolus on a un véritable cas d'incohérence qui suit directement de la définition du trait.

Beaucoup plus indirecte par rapport à l'existence d'une cohérence inter-situationnelle est la littérature sur les corrélations entre les traits (toujours mesurés par des questionnaires) et des conséquences générales telles que le succès au travail, dans les relations interpersonnelles, dans le mariage etc. Le problème est qu'on ne peut pas écarter l'hypothèse que bien d'autres facteurs interviennent et expliquent ces résultats : à supposer qu'il existe une honnêteté hyper cohérente, est-

³²⁰ Hugh Hartshorne, *Studies in the organization of character, Studies in the nature of character* ; v. 3 (New York, : Macmillan, 1930), Hugh Hartshorne et Mark Arthur May, *Studies in deceit, Studies in the nature of character* ; 1 (New York: Macmillan, 1928).

³²¹ En ce sens, cela est étonnant que, pour autant que nous sachions, les psychologues de la personnalité ne se soient pas attaqués à analyser minutieusement le comportement dans les « reality show », qui offrent une occasion unique.

ce qu'on doit s'attendre à beaucoup plus qu'une faible corrélation entre être honnête, ne pas se faire renvoyer du travail ou quitter par son partenaire ? Ce n'est pas évident, on peut imaginer que les honnêtes sont parfois exploités au travail ou par leur partenaire. Certes, pour une approche basée sur les traits une corrélation importante entre un trait et des accomplissements sera toujours une meilleure nouvelle qu'une corrélation faible. Cela prouve que le trait est très utile pour prédire le comportement, ce qui est une fin en soi pour les psychologues. Même dans ce cas de figure idéal dont personne ne se plaindrait, la question de la cohérence reste ouverte. La corrélation prouverait que le trait existe et a des conséquences très fiables, non pas qu'il est caractérisé par une attitude très cohérente. Il est possible qu'un niveau élevé de sociabilité soit seulement associé à des actes d'ouverture à autrui plus fréquents que la moyenne, et que toutefois la détection de cette attitude engendre des réactions d'une façon très fiable et régulière. Etre plus gentil que la moyenne (mais pas tout le temps) pourrait suffire pour se faire aimer par tout le monde. Ainsi, les conséquences des corrélations trait-accomplissement pour le problème de la cohérence sont plus difficiles à évaluer que celles d'une incohérence flagrante entre deux situations.

En restant dans le domaine de la psychologie de la personnalité, il semble qu'une étude systématique ou des données anecdotiques sur la biographie des personnes puissent donner beaucoup d'information sur le niveau de cohérence de leur attitude. Mais les problèmes d'interprétation ne manquent pas là non plus : on ne peut être sûr que le contexte de certaines actions précises soit rapporté de façon fidèle et les traits de caractère peuvent changer au fil des années. D'après les standards rigoureux de la psychologie expérimentale les données bibliographiques se révèlent souvent anecdotiques ; lorsqu'elles ne le sont pas, elles sont moins représentatives, car les études systématiques sont moins nombreuses. Le nombre de personnes pour lesquelles on dispose de données bibliographiques fiables est limité et l'échantillon n'est pas représentatif.

En général, le paradigme qui étudie la corrélation entre le comportement d'un groupe dans deux ou plusieurs situations spécifiques a été conçu comme le seul moyen pour prouver véritablement au-delà de tout doute raisonnable l'existence d'une cohérence inter-situationnelle. Pour ceux qui sont suffisamment convaincus par les données anecdotiques et la perception ordinaire de la cohérence, ce sera une preuve décisive. Pour les sceptiques, le seul terrain où montrer avec rigueur l'absence d'un phénomène fuyant mais qui pourrait se cacher derrière des liens trop indirects avec des accomplissements qui dépendent de trop de choses et ne sont pas adéquats pour le mesurer. En tout cas, tous voient dans ce genre d'études la seule possibilité de trancher la question. Que la solution soit déjà évidente ou non ailleurs, elle sera plus évidente ici.

Malgré cela, ces études restent plutôt rares : la raison principale est le coût et la difficulté d'organisation, par rapport au moyen commode du questionnaire sur lequel la psychologie de la personnalité investit majoritairement (ou des statistiques qui lient les questionnaires à des résultats objectifs facilement accessibles, comme le succès dans l'emploi). Par conséquent, l'étude de Hartshorne et May reste jusqu'à aujourd'hui la plus citée (ce qui n'est pas un bon signe, étant donné qu'elle date des années 20 !). Des milliers d'étudiants ont été testés dans plusieurs circonstances (chez eux, dans la classe, pendant des jeux avec leurs amis, dans des compétitions athlétiques etc.) dans lesquelles ils avaient l'opportunité de mentir, tricher et voler. Le genre de résultat qui a déclenché le débat a été que, alors que les étudiants se conduisaient de façon plutôt cohérente dans des situations spécifiques presque identiques (fonctionnellement identique : par exemple, copier sur la feuille des réponses d'un certain examen était corrélé 0,696 avec copier sur la feuille des réponses d'un autre examen) à des moments différents, montrant ainsi une considérable stabilité temporelle dans leur attitude, les corrélations tombaient progressivement avec une augmentation de la différence entre les deux situations³²². La corrélation entre copier sur la feuille des réponses et ajouter des points dans un test où les étudiants devaient reporter la vitesse de leur réponse était seulement 0,292. Pire, si on ajoute au changement de la tâche censée exprimer l'honnêteté un changement de l'environnement, les corrélations tombent davantage. La corrélation moyenne de quatre tests dans la classe était 0,292, celle des tests dans la classe avec des tests à l'extérieur 0,167. Par exemple, le test qui donnait l'opportunité de mentir dans la classe et celui qui permettait de tricher à l'extérieur avaient une corrélation d'à peine 0,064³²³. Globalement, la corrélation moyenne pour ces tests relatifs à l'honnêteté était de 0,26.

Dans une autre étude classique, Newcomb avait démontré une corrélation moyenne de 0,14 entre des situations pertinentes pour le trait de l'extraversion dans un groupe de 51 garçons avec des problèmes comportementaux³²⁴. Dudycha avait examiné la cohérence dans la ponctualité de 307 étudiants universitaires dans six situations (arriver en classe, à un rendez-vous, au petit déjeuner, à

³²² Le stabilité temporelle en réalité est un cas limité de cohérence inter-situationnelle dans lequel deux situations sont aussi proches que possible, la seule différence étant le moment de l'action. Par approximation, on parle de la même situation qui se répète, mais cela n'est pas vrai d'un point de vue rigoureux, car le fait que ce soit la première ou la millième fois qu'on se trouve dans un contexte peut bien sûr avoir une influence sur notre attitude.

³²³ Voir Walter Mischel, *Personality and assessment* (Mahwah, NJ: L. Erlbaum Associates, 1968) 24.

³²⁴ Theodore Mead Newcomb, "The consistency of certain extrovert-introvert behavior patterns in 51 problem boys" (Teachers College Columbia University, 1929).

l'église, etc.). La corrélation moyenne était de 0,19³²⁵. Dans une étude plus récente et méthodologiquement sophistiquée, Peake a suivi pendant dix semaines un groupe de 63 étudiants universitaires dans plusieurs situations censées être pertinentes pour le trait du « caractère amical » et de la « contrainte ». Corrélation moyenne : 0,08 pour la contrainte et 0,05 pour le caractère amical³²⁶.

Le nœud de la discussion est que des corrélations aussi faibles semblent incompatibles avec la diffusion dans la population d'une cohérence forte du comportement³²⁷. Avant d'analyser les réponses qui ont été apportées à ce problème par des auteurs qui reconnaissent son importance, il faut passer par deux tentatives pour montrer que le problème est illusoire : les données sur la corrélation des groupes ne donneraient aucune raison de mettre en question la cohérence inter-situationnelle des individus.

Funder : une tentative manquée de dépasser le « coefficient de personnalité »

Une première stratégie peu prometteuse consiste à négocier sur la valeur typique des corrélations dans les études qui ont adopté ce paradigme. 0,30 a été rendu célèbre par Mischel comme le « coefficient de personnalité », c'est-à-dire le plafond que les corrélations ne dépasseraient presque jamais³²⁸. Funder, citant un article du situationniste Nisbett, soutient que même les adversaires de l'approche des traits aujourd'hui concèdent un déplacement du plafond à 0,40³²⁹. Même si cela était correct, cela ne changerait pas grand chose, car le problème est posé par le niveau moyen des corrélations, non par le plafond. Que le plafond monte ne signifie rien s'il n'est pas suivi d'une augmentation de la moyenne. Or, Funder soutient aussi que la valeur moyenne des corrélations est beaucoup plus importante que les classiques en psychologie de la personnalité ne le montrent, cette fois sans citer un nouveau consensus dans le domaine (qui n'existe pas), mais en produisant

³²⁵ G. J. Dudycha, "An objective study of punctuality in relation to personality and achievement," *Archives of Psychology* 29 (1936).

³²⁶ P. K. Peake, "Searching for consistency: The Carleton student behavior study." (Thèse de doctorat, 1982).

³²⁷ On pourrait abandonner l'idéal d'une cohérence extrême : cette possibilité révisionniste est venue tout de suite à l'esprit aux psychologues. Comme nous le verrons, l'attitude révisionniste a une longue histoire : elle s'est imposée dans la pratique actuelle de la discipline et a influencé la théorie de la vertu de la psychologie positive, qu'on va analyser dans la suite. Cependant, les psychologues ont aussi cherché sans relâche la « super cohérence », qui fait partie des ambitions historiques de la théorie de la vertu. Ce chapitre est consacré uniquement à cette deuxième recherche.

³²⁸ Mischel, *Personality and assessment* 78.

³²⁹ Funder, "Personality," 199.

directement des données. Il croit pouvoir montrer que des corrélations de 0,40 sont la norme, et que souvent on peut atteindre des valeurs beaucoup plus importantes. S'il a raison, le problème qui a déclenché le débat sur la cohérence du comportement en psychologie de la personnalité ne se pose même pas. Malheureusement, son argument est fallacieux. Il sera très instructif de l'analyser dans les détails.

Funder part du constat correct que dans le débat personne-situation sur la cohérence du comportement on a eu tendance à mesurer la corrélation entre des comportements précis, par exemple, tricher dans le cadre d'un type d'examen et tricher dans un examen qui a un autre format. Sur cette base, il considère que les psychologues ont commis l'erreur de conceptualiser tous les traits comme ceux qu'on a appelé traits comportementaux, c'est-à-dire des traits qui prévoient toujours le même type d'action (« disposition à voler », « être souriant » etc.), non pas comme des traits plus complexes qui prévoient types d'actions différents selon le contexte. Il arrive même à ridiculiser l'attitude qui a amené à une erreur de ce genre :

« La conceptualisation usuelle de la cohérence du comportement était celle de quelque chose qui se manifeste au niveau des actions concrètes (...). Ce point de vue traite la mesure de la cohérence comme une simple question de calcul : si le nombre de fois qu'une personne fait quelque chose dans une situation est parfaitement prévisible à partir du nombre de fois qu'elle fait quelque chose d'équivalent dans une autre situation, alors son comportement est cohérent. Si ce nombre n'est pas prévisible de cette manière, alors son comportement est incohérent³³⁰. »

En pratique, il accuse les classiques d'avoir conceptualisé (par exemple) l'honnêteté comme « disposition à ne pas tricher » dont la cohérence est mesurable par un nombre constant ou variable d'épisodes de tricherie dans toutes les situations pertinentes. Cette conception serait bien sûr absurde, seul un psychopathe peut tricher tous le temps ou jamais, quoi qu'il arrive. Funder propose de prendre en compte la variabilité possible des comportements et de conceptualiser les traits comme des dispositions non-comportementales complexes. Cette critique est exactement la même qui a été avancée par les philosophes en réponse aux données expérimentales : vu que les expériences testaient la cohérence entre deux actions ou plus dans des contextes précis, c'est que le trait visé devait être conçu comme « disposition mécanique à faire C dans le contexte A et B ». A

³³⁰ D. C. Funder et C. R. Colvin, "EXPLORATIONS IN BEHAVIORAL CONSISTENCY - PROPERTIES OF PERSONS, SITUATIONS, AND BEHAVIORS," *Journal of Personality and Social Psychology* 60, no. 5 (1991): 774.

propos de ce genre d'expériences, Annas commente : « *La vertu aristotélicienne est conçue comme étant une habitude rigide et non critique*³³¹. »

La logique de Funder est totalement fallacieuse. Rien ne prouve que dans les études classiques sur la cohérence inter-situationnelle tous les traits étaient considérés comme comportementaux, et même si cela était le cas, les données resteraient incompatibles avec la cohérence de traits plus « intelligents » et complexes. Le fait qu'un trait soit complexe n'exclut pas qu'il puisse prescrire le même type d'action dans deux contextes différents. Imaginons une honnêteté basée sur un jugement raffiné, nuancé et une sagesse subtile. De quelqu'un d'honnête en ce sens on peut savoir qu'il ne va dérober ni son meilleur ami ni son colocataire (même type d'action : dérober), et si quelqu'un dérobe l'un ou l'autre, il n'est pas honnête, ni dans le sens riche ni dans le sens pauvre et comportemental. L'accusation de Funder est infondée. Ce serait comme si on accusait quelqu'un qui doute qu'un bon père puisse ne pas donner le même argent de poche à ses enfants du même âge (toute chose égale par ailleurs) de penser qu'être un bon père *est* « donner mécaniquement le même argent de poche à ses enfants ». En plus, même si les études classiques avaient mobilisé une conception comportementale des traits, dès que les situations testées se révèlent aussi (dans cette hypothèse, par coïncidence) diagnostiques de traits plus complexes, les données recueillies restent aussi bonnes et conclusives que dans le premier cas de figure.

Dans la même étude, Funder poursuit sa stratégie de décrédibilisation des traits comportementaux. Son paradigme est le suivant :

« 140 étudiants ont été observés dans deux situations expérimentales différentes. Dans la première, deux étudiants du sexe opposé qui ne s'étaient jamais rencontrés avant ont été introduits dans une petite pièce contenant peu de chose à part un canapé et une caméra. L'expérimentateur leur avait dit qu'ils pouvaient parler de ce qu'ils voulaient et qu'il reviendrait dans quelques minutes, après il avait activé la caméra (sous le regard des deux étudiants) et quitté la pièce. La deuxième situation avait lieu quelques semaines plus tard, et était exactement comme la première, sauf que chaque sujet avait un partenaire de sexe opposé différent, et les deux étaient là pour la deuxième fois.

*Le comportement typique dans cette situation était une espèce de conversation pour casser la glace, mais dans les limites de ce registre il y avait une large variation parmi les individus*³³². »

³³¹ Annas, "Comments on John Doris's 'Lack of Character'," 637.

³³² D. C. Funder, "Towards a resolution of the personality triad: Persons, situations, and behaviors," *Journal of Research in Personality* 40, no. 1 (2006): 23.

Les vidéos ont été évaluées et codées par 4 chercheurs sur une échelle de 1 à 9 pour 62 traits différents, dont par exemple « exprimer insécurité ou sensibilité », « tenir le partenaire à distance », « exprimer signes physiques de tension ou anxiété », « avoir l'air relaxé et à l'aise », « exhiber des compétences de socialisation », « sourire fréquemment », « parler fort » etc. Funder attribue à l'approche traditionnelle la prévision que, si le comportement doit être jugé cohérent, les participants devraient garder plus ou moins le même score absolu dans les deux situations. Par exemple, les étudiants devraient parler plus ou moins avec le même ton de voix dans les deux contextes, ou sourire exactement de la même façon. Les résultats, bien sûr, falsifient largement cette prévision. Le niveau absolu du ton de la voix (par exemple) varie entre les deux contextes. Funder considère que cela allait de soit, et qu'une prévision contraire ne fait que ridiculiser l'approche classique. Or, comme nous venons de le montrer, il est absolument injustifié d'attribuer une telle prévision à l'approche traditionnelle. La méthodologie raffinée de Harsthorne et May n'avait été mobilisée que pour découvrir que les gens ne gardent pas le même ton de voix ou ne sourient pas toujours de la même façon.

Cependant, les données de Funder semblent poser un vrai problème. Laissant de côté la question absurde du niveau absolu et mobilisant une définition acceptable, comparative et réaliste des traits concernés, qui puisse englober un niveau adéquat de variation situationnelle (« parler fort » deviendrait « avoir la tendance à parler plus fort que la moyenne dans une situation donnée »), Funder trouve des corrélations élevées pour la plupart des traits. Intuitivement, la différence fondamentale entre la première et deuxième situation est le niveau de familiarité des participants avec la tâche et le milieu du laboratoire : la deuxième fois ils sont en général plus détendus et en confiance, avec toutes les conséquences que cela entraîne. Il se trouve que leur attitude est souvent comparativement cohérente : par exemple, malgré la diminution (ou l'augmentation) du niveau absolu, ce sont toujours les mêmes qui ont un ton de voix ou une tendance à sourire plus élevés (ou inférieurs) que la moyenne. La plupart des corrélations trouvées atteignent le 0,40, et souvent elles sont beaucoup plus importantes. Si des données de ce genre sont acceptables, cela signifie qu'un problème de cohérence inter-situationnelle ne se pose même pas.

L'expérience présente au moins deux gros problèmes. D'abord, s'il est vrai que faire quelque chose la première fois et le refaire une deuxième fois peuvent être considérées comme deux situations différentes (la même chose, bien sûr, n'est pas valide pour une énième et une énième + 1 fois, autrement il n'y aurait pas de situations fonctionnellement équivalentes), il s'agit aussi de la plus petite différence qu'on puisse imaginer : c'est la différence qu'on ignore souvent dans la distinction entre cohérence inter-situationnelle et stabilité temporelle (censée désigner la répétition de la même situation, ce qui n'est jamais parfaitement exact). Nous avons vu que les corrélations décroissent

d'une façon proportionnelle à la différence des situations (qui sont néanmoins toutes diagnostiques du même trait), il n'est donc pas étonnant qu'elles soient plus importantes entre deux situations presque identiques. Cela est confirmé par d'autres aspects de l'expérience. Après la deuxième séance de conversation libre, les deux participants étaient invités à discuter pendant 5 minutes sur le thème de la peine de mort. L'expérimentateur leur concédait quelques minutes pour se préparer, mais les rôles de favorable ou opposé dans le débat étaient assignés au hasard, indépendamment des convictions personnelles. L'interaction était filmée et codée de la même manière que les deux précédentes. Le débat compte certainement comme une « autre situation » par rapport aux deux précédentes, mais, intuitivement, on peut s'attendre à ce qu'elle soit plus semblable à la deuxième situation qu'à la première. Même si les règles du jeu changent, le débat a lieu entre les mêmes couples de participants de la deuxième situation et dans le même climat de fond, vu qu'il suit immédiatement la discussion informelle. Le comportement des corrélations entre les différentes situations confirme cette attente :

« Entre les sessions 1 et 2, 37 (c'est-à-dire plus de la moitié) des 62 items ont atteint corrélations statistiquement significatives à un niveau de $p < 0,01$. Le nombre d'items atteignant ce niveau de cohérence entre la session 2 et le débat était 26 ; le nombre entre la session 1 et le débat était 18. Presque toutes ces corrélations ont été répliquées indépendamment dans les sous-groupes de filles et garçons (...).

Globalement, 61 sur 62 corrélations inter-situationnelles entre la session 1 et 2 étaient positives, et 45 étaient statistiquement significatives ($p < 0,5$). Entre la session 2 et le débat, 57 corrélations étaient positives, et 41 étaient statistiquement significatives ($p < 0,5$) ; entre la session 1 et le débat, 56 corrélations étaient positives, et 30 étaient statistiquement significatives ($p < 0,5$). »

On voit que dans le cas le plus difficile (corrélations entre la session 1 et le débat) les résultats sont plus modestes. En effet, dans ce cas il n'y a que 11 corrélations (sur 62) supérieures ou égales à 0,38³³³. Cela pourrait quand même être très suggestif et prometteur : 11 traits dont on a prouvé un niveau de cohérence très intéressant dans deux situations vraiment différentes c'est déjà beaucoup face à une tradition très pessimiste. Cela pourrait être le début d'un programme de recherche qui en découvre beaucoup plus.

Malheureusement, il y a un deuxième problème. Funder a sans doute choisi de se faciliter la tâche en étudiant des traits de très bas niveau, presque des réflexes. En tout cas, les 11 traits qui sont

³³³ Funder et Colvin, "EXPLORATIONS IN BEHAVIORAL CONSISTENCY - PROPERTIES OF PERSONS, SITUATIONS, AND BEHAVIORS," 782.

associés à un niveau de cohérence intéressant sont de ce genre, par exemple : « parler fort » (0,58), « parler rapidement » (0,56), « être expressif dans le visage, la voix, les gestes » (0,52), « chercher constamment le contact visuel avec le partenaire » (0,42) etc. D'après notre deuxième principe méthodologique, une cohérence au niveau de réflexes et d'attitudes de base dans la façon de parler et de bouger ne nous dit rien sur la cohérence de traits plus complexes, car il s'agit de processus psychologiques complètement différents. Avoir un positionnement plutôt stable du niveau de sa propre voix par rapport à la moyenne paraît quelque chose qui relève presque du physiologique, rien à voir avec le fait d'être constamment généreux ou courageux. En conclusion, la tentative de Funder de montrer que le débat sur la cohérence du comportement a été stimulé par des mauvaises données échoue. D'après les résultats des expériences pertinentes, les corrélations du comportement des gens entre plusieurs situations sont en effet en moyenne très peu élevées, de l'ordre du 0,20 au 0,30, voire beaucoup moins. Funder n'a montré que la cohérence possible de simples réflexes de bas niveau.

Vranas : le plafond de 0,30 est réel mais compatible avec un niveau élevé de cohérence

Depuis les années 20, le débat a été déclenché par le constat que ces bas coefficients posent pour la cohérence inter-situationnelle un problème auquel il faut trouver une solution. Récemment, Vranas a rejoint Funder dans son idée que le débat n'aurait dû avoir lieu, mais non pas parce que les coefficients seraient plus élevés que ce qu'on a cru. D'après Vranas, pour une classe importante de traits, de faibles coefficients de ce genre ne sont simplement pas liés à la cohérence inter-situationnelle. Les gens pourraient être cohérents dans leur attitude alors même que les corrélations dans les expériences seraient peu élevées³³⁴. Son argument est ingénieux. Les corrélations sont directement liées uniquement à la cohérence comparative, c'est à dire au positionnement constant par rapport au niveau moyen (ex : parler plus fort que les autres). Si elles sont peu élevées, c'est forcément que ce positionnement relatif n'est pas constant (ex : x n'est pas toujours celui qui parle plus fort que les autres)³³⁵. Or, une classe importante de traits parmi les plus significatifs, dont les vertus morales, n'est pas définie par une cohérence comparative, mais par une cohérence de niveau absolu. Pour s'en rendre compte, il suffit d'imaginer une population où tout le monde fait quelque

³³⁴ Pour une autre stratégie qui défend la compatibilité entre le plafond de 0,30 et l'existence de traits larges et cohérents – celle de Sabini – voir la section sur les méta-traits ci-dessous.

³³⁵ Pour la preuve statistique, voir Peter B. M. Vranas, "Moral Behavior: Cross-Situational Consistency and Predictability" (non publié: The University of Michigan, 2006).

chose d'admirable dans une situation (ex : sauver un enfant qui est en train de se noyer) et quelque chose de condamnable dans une autre situation (ex : voler de l'argent à un ami). D'un point de vue comparatif, un individu qui fait cela (comme tous les autres) serait cohérent, alors qu'il est évident qu'il s'agit d'une incohérence si on regarde au niveau absolu de moralité, et c'est bien cette deuxième perspective qui est importante.

Le dernier exemple montre qu'une cohérence comparative élevée est compatible avec une incohérence de niveau absolu. Ce cas de figure ne nous intéresse pas, car dans les expériences classiques la cohérence comparative (mesurée par les corrélations) est faible. Est-il possible d'avoir une cohérence comparative faible avec une cohérence absolue élevée ? Vranas soutient que oui. Il donne un exemple hypothétique :

« Dans le test 1, 2, et 3, l'étudiant 1 a un score respectivement de 90, 88, 89 ; l'étudiant 2 de 89, 90, 88 ; et l'étudiant 3 de 88,89, et 90. Chaque étudiant montre une cohérence non-comparative (il a plus ou moins le même score dans tous les trois tests) mais n'affiche pas de cohérence comparative (pour exemple, l'étudiant 1 est le premier dans le test 1, le dernier dans le test 2, au milieu dans le test 3)³³⁶. »

Je ne suis pas convaincu. Si la cohérence non-comparative est élevée, c'est que chaque étudiant a plus au moins toujours la même note, par hypothèse. Mais alors, à moins de supposer que tous ont plus au moins le même niveau (comme dans l'exemple), il est difficile de voir comment la cohérence comparative pourrait ne pas être élevée aussi. S'il y a des bons, des mauvais et des moyens et que tout le monde a plus ou moins toujours la même note, la hiérarchie va rester bien stable. Nous ne voyons pas pourquoi nous devrions accepter la condition selon laquelle tous ont en gros le même niveau, il s'agit d'une hypothèse qui paraît particulièrement absurde pour la morale (tous sont-ils soit bons soit mauvais, ou neutres ?) et d'ailleurs pour n'importe quel trait (la notion de trait en psychologie et dans le sens commun est censée rendre compte de la variation individuelle...). Nous ne voyons pas d'autres hypothèses plausibles qui pourraient rendre compatible un niveau élevé de cohérence non-comparative avec un niveau faible de cohérence comparative. Les corrélations faibles représentent bien un problème pour la cohérence inter-situationnelle non-comparative.

A la recherche de la cohérence perdue

³³⁶ Peter B. M. Vranas, "Lack of character: Personality and moral behavior," *Philosophical Review* 113, no. 2 (2004): 286.

Que faire face à des corrélations aussi faibles ? Nous avons vu qu'elles ne sont pas un artefact expérimental (contre Funder) et que, *prima facie*, elles posent bien un problème pour la cohérence inter-situationnelle (contre Vranas). Les psychologues de la personnalité ont cherché à répondre à ce problème en explorant une série d'hypothèses alternatives. En fait, l'espace logique pour rendre compatibles les résultats de la plupart de ces expériences avec une conception forte de cohérence comportementale est vaste. Les possibilités qu'on pourrait explorer sont les suivantes :

traits objectifs bizarres ou non représentés dans le langage : on a la mauvaise liste (inspirée par le vocabulaire ordinaire) des traits objectifs censés exister. Il y en a, mais ils sont différents du courage, de l'extraversion, de la timidité etc.

perception différente mais partagée des situations : on a une mauvaise théorie de la façon dont les gens catégorisent les situations et l'équivalence fonctionnelle entre elles. A cela près, il y aurait bien des traits objectifs qui rassemblent beaucoup à la conception traditionnelle

nature subjective de la plupart des traits : il y aurait une perception différente des situations dans des individus différents, la plupart des traits qui existent sont subjectifs.

méta-traits : un bon niveau de cohérence pourraient être limité à certains traits pour un certain type de personne : seulement certaines personnes auraient pour ainsi dire le « méta-trait » d'avoir des traits.

Existence de traits non intuitifs ou non représentés dans le langage

Pour comprendre la première hypothèse, il suffit de penser à une comparaison interculturelle. Il est vraisemblable que dans d'autres cultures des comportements spécifiques associés à la vertu du courage changent. Imaginons par exemple une culture où les adolescents doivent être laissés se débrouiller le plus possible, sauf cas extrêmes de danger. Si on allait tester la cohérence de l'attitude des membres de cette culture entre deux situations, où il s'agit d'aider des adolescents dans un cas, et des adultes dans l'autre, les résultats seraient décevants. Mais cela parce qu'on a testé le mauvais trait, il aurait fallu tester le courage dans la bonne définition locale. Or, la première hypothèse imagine que ce type de problèmes pourraient arriver même chez nous. Il est possible que des traits objectifs et très cohérents existent, mais qu'on ait de mauvaises attentes. Il faudrait chercher des traits objectifs définis de façon différente et qui pourraient ne correspondre pas à notre liste ordinaire. Cette hypothèse n'a pas été spécialement explorée par les psychologues des traits dans le sens de variations « ad hoc » et arbitraires de traits du sens commun. Par exemple, on n'a pas testé l'hypothèse que le courage réellement existant s'éloigne de la signification communément partagée en ce qu'on ne va pas aider des adolescents (sauf urgence extrême). De même, on n'a pas testé

l'existence de l'*airpassion* dont parle Doris en tant que version hétérodoxe de la compassion³³⁷. Cela est bien compréhensible : on ne voit pas pourquoi les psychologues devraient procéder d'une telle façon arbitraire. Par contre, bien plus intéressante est l'hypothèse qu'il y ait des traits importants qui ne trouvent pas une définition directe dans le langage. Cette voie a été explorée abondamment, mais sans le résultat d'augmenter les corrélations situationnelles moyennes. La littérature en psychologie de la personnalité abonde en traits absents du langage commun et définis de façon stipulative. Comme le souligne Funder : « *Plusieurs variables contemporaines de personnalité sont relativement ésotériques – elles sont délibérément non intuitives ou même contre-intuitives*³³⁸. » Par exemple, Cattell a puisé dans la psychologie clinique en introduisant des termes relatifs aux troubles de l'humeur tels que cyclothymie et schizothymie, conceptualisés en tant que traits³³⁹. Rien de cela n'a aidé à atteindre de coefficients de corrélations plus importants.

La perception partagée des situations ne correspond pas à nos attentes

La deuxième hypothèse a également fait l'objet d'études. Il s'agit d'une idée subtile et ingénieuse. Avant de jeter à la mer trop hâtivement l'idée traditionnelle de trait objectif, où tout le monde perçoit la situation de la même façon et réagit de la même façon (au moins dans des cas prototypiques qui définissent l'essence d'un trait), on peut penser que les gens catégorisent bien de la même manière les situations cruciales, mais d'une façon qui n'est pas celle décrite par nos théories psychologiques. Cela est compatible avec l'existence des traits du sens commun : ils se manifesteraient dans des situations différentes de celles anticipées par les expérimentateurs. Par exemples, dans les expériences de Harsthorne et May les étudiants trichaient dans un contexte et beaucoup moins dans d'autres. Il est possible qu'ils ne voyaient pas les deux circonstances comme étant équivalentes : peut-être à tout le monde il paraissait évident qu'ajouter quelques points dans une tâche de vitesse auto-évaluée est beaucoup moins grave que copier délibérément sur la feuille

³³⁷ L'*airpassion* est une forme fictionnelle de compassion qui ne serait plus valable à des altitudes élevées : par exemple, on aurait le droit de ne pas abandonner son escalade pour aller au secours d'un autre alpiniste en détresse. Doris introduit cette notion fictionnelle pour montrer que, en postulant des traits *ad hoc*, il est possible de préserver la cohérence du comportement : un alpiniste qui aide les personnes en détresse à la maison mais pas à la montagne est cohérent, du point de vue de l'*airpassion*. Doris, *Lack of character: personality and moral behavior* 80.

³³⁸ D. C. Funder, "GLOBAL TRAITS - A NEO-ALLPORTIAN APPROACH TO PERSONALITY," *Psychological Science* 2, no. 1 (1991): 31.

³³⁹ R. B. Cattell, "The description of personality: Basic traits resolved into clusters," *Journal of Abnormal and Social Psychology* 38, no. 4 (1943).

des réponses. Au contraire, il est possible que copier directement les réponses ait été perçu comme un « devoir moral » pour faire circuler l'information dans la classe.

Ce genre d'hypothèses ne vont pas satisfaire celui qui, adossé à ses choix normatifs, impose à l'honnêteté d'être incompatible avec toute forme de tricherie. Mais, à ce stade, nous nous intéressons uniquement à l'existence d'une conduite cohérente en tant que telle : si l'on pouvait prouver l'existence d'une forme d'« honnêteté » un peu déviante (par rapport aux attentes des philosophes et des psychologues) mais extrêmement cohérente dans le comportement des gens, ce serait un résultat énorme. L'hypothèse est également beaucoup plus facilement testable que l'existence de traits subjectifs, où les sceptiques voient le risque de postuler que chacun a « son honnêteté à lui », car cette perception et catégorisation alternative des situations qui échapperait aux psychologues est censée être universellement partagée.

Autrement dit, il nous manque une théorie générale de la perception des situations sociales. Si on savait ce que les gens perçoivent comme saillant, on pourrait mettre au jour la cohérence du comportement humain. Malheureusement, une théorie générale de la perception des situations – depuis plusieurs décennies considérée comme nécessaire et fondamentale dans le milieu de la psychologie de la personnalité – n'en est aujourd'hui qu'à ses premiers balbutiements, loin en tout cas de l'état où elle pourrait être utile pour évaluer sérieusement cette hypothèse. De quoi s'agirait-il concrètement en tout cas ?

Il s'agirait d'une taxonomie générale des situations sociales. Si l'idée paraît absurde et trop ambitieuse, étant donné la subtilité des différentes circonstances qui se succèdent dans la vie de tous les jours, c'est probablement parce qu'on a des attentes théoriques exagérées : il ne s'agit pas de classer toute situation envisageable comme on classerait toute espèce en biologie. Il s'agirait de déceler un certain nombre de lois, de biais et d'effets généraux portant sur la façon dont les êtres humains catégorisent les situations sociales. Ce serait l'équivalent, pour la perception sociale, de l'étude des lois et des biais de base de la perception visuelle. Cela nous permettrait de découvrir les facteurs principaux qui font que deux situations sociales soient traitées comme similaires par le système cognitif. Si l'on disposait d'une théorie de ce genre, on pourrait prédire systématiquement des effets tels que (par exemple) la différence de perception entre le vol d'un cd dans un magasin et le téléchargement illégal en ligne³⁴⁰.

³⁴⁰ L'existence d'un ensemble de « lois » de base de la catégorisation des situations sociales n'est pas prouvée par le phénomène de la variation dans la perception sociale. Cette dernière pourrait être expliquée par une différence au niveau de croyances générales. Par exemple, on pourrait avoir une conception générale du rôle d'Internet dans la société qui est compatible avec la pratique du téléchargement illimité. L'existence de lois générales de la perception sociale constitue bien une hypothèse ambitieuse.

Comme nous l'avons déjà anticipé, aucune classification n'a atteint un niveau d'acceptation et de pouvoir prédictif intéressant. Dans le cadre de la psychologie de la personnalité³⁴¹, les tentatives pour réaliser un tel projet théorique ne sont que récentes : nous allons les passer rapidement en revue. Ce sera clair qu'aucune n'est en mesure de jouer un rôle dans un argument qui prouverait l'existence de la cohérence inter-situationnelle dans le comportement.

Découverte des dimensions générales qui expliquent la similitude entre situations
Une approche pertinente pour l'étude de la cohérence dans le comportement consiste à inférer de récits ou de descriptions spontanées (ou sollicitées directement par l'expérimentateur) le genre de situations qui sont décisives pour la manifestation d'un ou plusieurs traits donnés. Dans les expériences censées tester la cohérence du comportement, les intuitions théoriques des expérimentateurs qui guident le choix des situations « test » (entre lesquelles la cohérence devrait se manifester, p.ex. honnête dans le cours de récréation, honnête pendant l'examen final) pourraient ne pas correspondre à l'importance et au niveau de similitude perçus par les sujets. Si on se limitait à prendre en considération uniquement les situations considérées comme importantes par les sujets, leur comportement pourrait se révéler cohérent (dans ces situations).

Lorsque les gens se plaignent du manque de courage ambiant ou louent l'honnêteté, quels exemples citent-ils de préférence ? En se posant des questions de ce type, Ten Berge et De Raad « ont demandé aux participants de décrire des situations dans lesquelles des personnes avaient manifesté des traits de personnalité spécifiques, générant ainsi une liste de situations qui sont utiles pour décrire des différences individuelles³⁴². » Le résultat a été une liste de 237 situations. Après quoi :

« Nous avons classifié les situations en demandant aux participants d'indiquer dans quelles mesure ils se sentiraient capable d'y faire face. Selon leur structure psychologique, on s'attend à ce que les gens soient différemment à l'aise dans certaines situations. Cette étude décrit la construction d'une taxinomie générale des situations basée sur des évaluations (par un sujet ou par un tiers) de "se débrouiller dans une situation donnée"³⁴³. »

³⁴¹ Bien évidemment, l'énorme littérature sur les biais dans la cognition sociale serait pertinente ici (certains de ces biais opèrent vraisemblablement au niveau de la perception), mais aucun psychologue de la personnalité ne l'a résumée et mobilisée de façon systématique pour la mettre au service de la stratégie d'étude de la cohérence du comportement que nous sommes en train d'analyser.

³⁴² M. A. Ten Berge et B. De Raad, "The structure of situations from a personality perspective," *European Journal of Personality* 16, no. 2 (2002): 83.

³⁴³ Ibid.

L'analyse factorielle a donné une solution avec quatre facteurs : plaisir, difficulté individuelle, conflit interpersonnel, pression sociale. Chaque situation de la liste est associée à une certaine valeur des quatre facteurs. Autrement dit, les expérimentateurs ont déniché des facteurs généraux qui expliquent le niveau de « similitude perçue » des situations (dans ce cas, en ce qui concerne le tendance à se sentir à l'aise). Deux situations qui sont associées à une valeur élevée de « pression sociale » seront perçues comme étant deux situations du même type (fonctionnellement équivalentes). On ressent probablement des exigences de cohérence beaucoup plus fortes à l'intérieur de ces classes d'équivalence (ensemble de situations associées à peu près au même score des facteurs généraux) qu'entre des situations qui appartiennent à des classes d'équivalence différentes. Par exemple, on serait plus cohérent dans différentes situations de fort conflit interpersonnel qu'entre des situations de pression sociale et de conflit interpersonnel. C'est là l'idée de base d'une stratégie ingénieuse pour prouver la cohérence du comportement. Ce serait passionnant d'avoir une vision d'ensemble des dimensions pertinentes pour évaluer la similitude perçue des situations, mais, malheureusement, ce genre de résultat n'a pas encore été produit : l'étude de cette stratégie n'en est qu'à ses balbutiements.

Template Matching

Une stratégie différente pour en savoir plus sur les équivalences fonctionnelles perçues par les gens est celle du *Template Matching*³⁴⁴. L'idée est d'évaluer le niveau de similitude entre deux situations en comparant le profil typique de ceux qui se comportent de la même façon dans les deux :

« Considérez deux universités, UC Berkeley et Harvey Mudd College. Le genre d'étudiant qui va réussir à Berkeley est décidé, sociable et motivé par soi-même. Un étudiant qui se met moins en avant, est moins engagé dans le social et a besoin d'un apport plus constant de conseils et de soutien de l'extérieur va avoir plus de chance de s'épanouir à Harvey Mudd. Remarquez que rien n'a encore été dit sur les différences dans la politique d'admission des étudiants, l'ambiance ou la quantité de soutien fourni par les enseignants dans ces deux institutions, mais une idée claire de leur différence commence déjà à émerger³⁴⁵. »

L'intuition est que si le même style de gens réussissent ou échouent, c'est que les deux situations sont psychologiquement très proches. Inversement, si dans deux situations qu'on considère

³⁴⁴ D. J. Bem et D. C. Funder, "PREDICTING MORE OF PEOPLE MORE OF TIME - ASSESSING PERSONALITY OF SITUATIONS," *Psychological Review* 85, no. 6 (1978).

³⁴⁵ Ibid.: 30.

intuitivement très semblables les profils de ceux qui réussissent (ou échouent) sont différents, c'est que les deux situations sont en réalité différentes d'une façon psychologiquement significative. Si le profil de l'étudiant qui est le plus typiquement honnête dans la cour de récréation ne coïncide pas avec le profil de l'étudiant qui est le plus typiquement honnête pendant l'examen final, cela signifie que les deux situations (avoir la possibilité de tricher pendant un jeu dans la cour/ avoir la possibilité de tricher pendant l'examen final) ne sont pas fonctionnellement équivalentes, en ce qui concerne l'honnêteté. Le *template matching* est une méthode pour découvrir les véritables rapports de similitude entre situations dans la perception sociale.

Comme on peut l'imaginer, une fois redéfinis les liens de similitudes des situations sociales par cette méthode, on saura où chercher et tester la cohérence du comportement : entre des situations qui soient vraiment semblables et fonctionnellement équivalentes.

Dans l'article cité, Bem et Funder croient pouvoir appliquer la méthode dans un cas de similitude apparente, qui cacherait une différence profonde entre deux situations. Avec 29 enfants, ils répliquent de façon plutôt fidèle un paradigme expérimental classique sur le « délai de gratification »³⁴⁶. Les enfants peuvent choisir entre attendre une quinzaine de minutes pour avoir deux bonbons ou activer une sonnette et avoir tout de suite un seul bonbon (la réplique n'est pas entièrement fidèle car dans le paradigme original l'expérimentateur sort pendant l'attente, alors que dans l'expérience de Funder il reste sur place). En général, ce paradigme s'est révélé extrêmement fructueux en faisant l'objet de plusieurs décennies de recherche : le fait d'être en mesure d'attendre pour avoir deux bonbons à l'âge de 4 ans a une très grande capacité prédictive par rapport à des accomplissements (par exemple scolaires) 10 (voire plus) années plus tard, il s'agit même d'une des variables personnelles qui ont le plus de pouvoir prédictif dans la littérature.

Dans ce contexte, le but de Funder était de montrer que deux situations apparemment proches (l'attente de l'enfant dans l'expérience originale de Mischel et dans cette réplique) pouvaient cacher de subtiles différences psychologiques : grâce à la méthode du « *template matching* », il découvre que le profil typique de l'enfant en mesure d'attendre dans les deux cas n'est pas le même. En particulier, il paraît que l'enfant patient n'est pas spécialement intelligent, loquace, désireux d'apprendre ou ouvert à des nouvelles expériences, alors que dans le paradigme classique le résultat était exactement le contraire. Conclusion : il y a une différence fonctionnelle cachée entre les deux situations (l'attente de l'enfant dans l'expérience originale de Mischel, et dans cette réplique).

³⁴⁶ W. Mischel, A. R. Zeiss et E. B. Ebbesen, "Cognitive And Attentional Mechanism In Delay Of Gratification," *Journal of Personality and Social Psychology* 21, no. 2 (1972). Pour l'explication du paradigme expérimental, voir le chapitre sur les approches socio-cognitives, paragraphe RS et auto-régulation.

Il faut remarquer que cette expérience n'articule que la moitié de la démonstration requise pour réhabiliter l'idée d'une cohérence comportementale forte : après avoir redéfini de manière informative les liens de similitude entre des situations, il faut montrer expérimentalement que le comportement est très cohérent dans plusieurs situations qui sont véritablement proches. Cette deuxième phase n'a pas été explorée du tout dans la littérature. Malheureusement, pour la stratégie de « template matching », la situation est encore plus négative. Le résultat de Funder, avec son profil typique des enfants patients, n'a pas été répliqué par une tentative explicite de Mischel, qui a retrouvé l'image habituelle d'un enfant intelligent, loquace et ouvert à de nouvelles expériences³⁴⁷. Pire encore, la stratégie de base a des précédents négatifs. Comme Mischel le montre, la méthode partage des présupposés avec une approche classique en psychologie de la personnalité : l'élaboration de « cookbooks » ou « atlas », rendue célèbre par le psychologue Paul Meehl³⁴⁸.

L'idée du *cookbook* est la suivante : pour prédire comment quelqu'un va se comporter dans une certaine situation ou dans un certain rôle, il faut d'abord disposer d'une base de données qui détaille chacun des profils typiques associés à différents types d'action et de situations. Par exemple, le profil typique d'un braqueur de banque, d'un professeur qui publie dans les meilleures revues, de quelqu'un qui va essayer de se suicider, etc. Ensuite, pour prédire le comportement d'une personne donnée, il s'agit d'établir combien elle s'éloigne de ce profil idéal : la différence dans le comportement sera proportionnelle. On va d'autant plus s'attendre à que vous fassiez x que vous vous rapprochez davantage du profil typique de la personne qui fait x ³⁴⁹.

On voit que le « *template matching* » partage avec le « *cookbook* » l'idée qu'il existe une « personnalité de la situation », un profil typique et stable qui de quelqu'un qui fait x dans la situation concernée. Le problème des deux approches est le même : ce genre de profils de la situation se révèlent finalement ne pas être stables. Leur stabilité locale peut paraître impressionnante et prometteuse : par exemple, il ne serait pas étonnant si on pouvait montrer qu'il y a un profil bien précis du policier indiscipliné (susceptible de violer la procédure dans une série de situations x, y, z) dans le San Francisco des années 1980, où du névrosé dans le Paris des années 1950. Par contre, l'utilité de ces généralisations disparaîtrait immédiatement si on changeait de lieux ou de période. Les *cookbooks* sont toujours locaux et il faut en faire constamment de

³⁴⁷ Walter Mischel et Philip K. Peake, "Beyond Deja-Vu in the Search for Cross-Situational Consistency," *Psychological Review* 89, no. 6 (1982).

³⁴⁸ P. E. Meehl, "Wanted—A good cookbook," *American Psychologist* 11 (1956).

³⁴⁹ *Cookbook* parce que, comme un livre de recettes, la base de données nous informe sur les « ingrédients » nécessaires (les caractéristiques du profil) pour obtenir le résultat voulu (l'action correspondante, p.ex. braquer une banque).

nouveaux³⁵⁰. L'exemple du manque de répliation par Mischel du résultat de Funder ne promet rien de meilleur. Le *template matching* assume (tout comme le *cookbook*) la stabilité de la personnalité typique associée à la situation : étant donné que le profil typique change, ce n'est pas possible de redéfinir les liens de similitudes et d'équivalence fonctionnelle par cette méthode. Par conséquent, le *template matching* ne peut pas être utilisé pour prouver la cohérence du comportement.

En conclusion, l'idée d'étudier la perception et la catégorisation des situations sociales, dans l'espoir de découvrir des universaux dans la catégorisation, n'a pas encore donné des résultats pertinents pour le débat sur la cohérence inter-situationnelle.

Nature subjective des traits

D'après la troisième hypothèse, des traits de caractères cohérents existent bien, mais ils sont subjectifs. Comme nous l'avons déjà expliqué, cela ne veut pas dire que n'importe quelle attitude vers des situations dangereuses pourra être appelée « mon courage à moi », en rendant la notion tautologique et vide. Au contraire, il s'agit de l'idée que tout le monde réagit de façon cohérente et homogène, lorsqu'il interprète la situation de la même façon, ce qui n'est pourtant pas toujours le cas. Un exemple peut être l'irascibilité définie comme réaction violente lorsqu'on se sent menacé, où la perception de la menace peut être très subjective.

En psychologie de la personnalité, ce genre d'hypothèse a séduit beaucoup plus les chercheurs occupés à élaborer un cadre conceptuel d'arrière-plan que les psychologues expérimentaux. En ce qui concerne le premier phénomène, on peut remonter jusqu'à Allport, le véritable père de la

³⁵⁰ Les travaux de Meehl en psychologie de la personnalité ont donné naissance à la tradition des formules de prédiction statistique, par le biais desquelles le comportement ou la performance sont prédits à partir d'un nombre limité de données (ex : niveau scolaire, résultat dans un test d'admission) dans une simple formule linéaire $ak+bk^2=c$ (où k et k^2 sont des coefficients), voir Bishop et Trout, *Epistemology and the psychology of human judgment*. Le lien historique entre les deux sujets est très intéressant : les « cookbooks » ont été développés pour montrer l'utilité des tests de personnalité (car les traits de personnalités étaient mobilisés pour définir les profils), mais on s'est vite rendu compte que, si l'on cherche des corrélats personnels de la prestation dans une situation spécifique, il vaut mieux utiliser d'autres facteurs plus simples et faciles à mesurer. Pourquoi s'attarderait-on à montrer qu'une personne honnête, généreuse et courageuse va typiquement obtenir une promotion, si en prenant en compte le niveau scolaire ou le salaire on peut formuler des prévisions aussi bonnes ? Autrement dit, formuler les profils du *cookbook* en termes de traits de personnalité est un effort inutile : il vaut mieux définir les profils par des facteurs les plus simples possibles. De stratégie prometteuse dans le cadre de la psychologie de la personnalité, l'approche inaugurée par Meehl est devenue l'une des traditions de recherche les plus critiques envers ce domaine.

discipline. Depuis 1937 il soulignait que, entre des situations, « *la similarité est personnelle* ³⁵¹. » Plus récemment, Bem et Allen ont soutenu que « *la classification des situations doit être une partie intégrante de l'évaluation ; en plus...cette classification doit être faite dans les termes de la phénoménologie de l'individu, non de l'expérimentateur* ³⁵². » Mischel semble être d'accord : « *clairement, des personnes différentes peuvent regrouper et coder les même événements de façon différente(...un stimulus perçu comme dangereux et menaçant par un individu peut être vu comme un défi ou comme étant excitant par son voisin* ³⁵³. »

Cependant, il semble que cet enthousiasme théorique s'est vite estompé face à la difficulté de construire des paradigmes expérimentaux dans lesquels on peut tester indépendamment la perception du contexte à l'échelle individuelle *et* le comportement. Cet objectif est certainement réalisable, car il y a un bon nombre de corrélats psychophysiologiques à la catégorisation subjective d'une situation (ex : transpiration, réactions physiologiques spécifiques liées à des émotions différentes) qui peuvent être contrôlés, mais cela n'est sans doute ni pratique ni économique. Ainsi, cédant à la tentation de paradigmes expérimentaux plus gérables, il y a eu plusieurs expériences qui ont contrôlé indépendamment la perception de similitude entre plusieurs situations et comment les gens *prévoient* qu'ils réagiraient dans ces situations ³⁵⁴. Ces résultats paraissent réconfortants pour l'hypothèse de traits subjectifs, car lorsque quelqu'un voit de façon idiosyncrasique deux situations comme étant plutôt différentes, il prévoit s'y conduire aussi de façon différente. Mais, bien sûr, on ne peut pas se contenter de ce que les gens disent qu'ils feraient, surtout dans ce genre de débat qui met en question les attentes et la perception ordinaire des gens.

Parmi le nombre limité de bonnes expériences sur ce problème, celle de Lord mérite une mention particulière ³⁵⁵. Il a évalué la capacité prédictive de la perception subjective des situations par rapport au comportement de 40 sujets dans six situations objectives différentes censées être liées au

³⁵¹ Gordon W. Allport, *Personality; a psychological interpretation* (New York,: H. Holt and Company, 1937) 283.
Italique dans le texte.

³⁵² D. J. Bem et A. Allen, "ON PREDICTING SOME OF PEOPLE SOME OF TIME - SEARCH FOR CROSS-SITUATIONAL CONSISTENCIES IN BEHAVIOR," *Psychological Review* 81, no. 6 (1974): 518.

³⁵³ Walter Mischel, "The interaction of person and situation," in *Personality at the crossroads: Current issues in interactional psychology*, édition S. Endler D. Magnusson (Hillsdale, NJ: Erlbaum, 1977), 342.

³⁵⁴ E. G. Klirs et W. Revelle, "PREDICTING VARIABILITY FROM PERCEIVED SITUATIONAL SIMILARITY," *Journal of Research in Personality* 20, no. 1 (1986), Barbara Krahé, *Situation cognition and coherence in personality : an individual-centered approach, European monographs in social psychology* (Cambridge England ; New York Paris: Cambridge University Press ; Editions de la Maison des Sciences de l'Homme, 1990).

³⁵⁵ C. G. Lord, "PREDICTING BEHAVIORAL CONSISTENCY FROM AN INDIVIDUALS PERCEPTION OF SITUATIONAL SIMILARITIES," *Journal of Personality and Social Psychology* 42, no. 6 (1982).

caractère consciencieux (par exemple, « prendre ses notes soigneusement » ou « s’habiller proprement »). Non seulement le nombre de situations où le comportement effectif a été mesuré (six) est remarquable pour les standards de cette littérature, l’auteur a aussi utilisé quatre méthodes différentes et non équivalentes pour mesurer la perception subjective : la première consistait simplement à demander d’estimer la similitude sur une échelle de sept points pour tout couple de situations liées au caractère consciencieux.

La seconde essayait d’arriver à une classification des situations selon l’utilité instrumentale de ces dernières pour atteindre des buts considérés importants par l’étudiant (par exemple, il est possible que prendre soigneusement des notes et s’habiller correctement soient psychologiquement associés dans l’esprit d’un individu au succès scolaire, alors que tenir en ordre sa propre chambre n’a rien à voir avec cela ; les deux premières situations et non pas la troisième feraient partie de la même classe d’équivalence, ce qui pourrait expliquer les niveaux de cohérence différents).

La troisième méthode, inspirée de la stratégie du *template matching*, consiste à demander à chaque sujet de construire un profil typique de la personne qui serait la plus consciencieuse imaginable dans chacune des six situations (six profils différents, il ne s’agit pas de la personne qui serait la plus consciencieuse dans toutes). Après, chacun doit aussi construire un profil pour lui-même : deux situations seront considérées comme étant proches pour un individu donné si leur deux profils idéaux (faits toujours par le même individu) différent à peu près au même niveau du profil du sujet. Ainsi, si d’après x son profil n’est trop différent ni de quelqu’un qui s’habille d’une façon parfaitement correcte à la perfection ni de quelqu’un qui prend toujours soigneusement ses notes, c’est que x considère que les deux situations appartiennent à la même classe d’équivalence : son comportement devrait être très cohérent entre les deux.

La quatrième méthode est une variante plus simple de la troisième : on prend les six profils idéaux faits par x et on considère que pour lui deux situations sont proches si leur deux profils idéaux se rassemblent, éloignées s’ils sont très différents. Ainsi, si d’après x le profil de quelqu’un qui s’habille d’une façon parfaitement correcte et de quelqu’un qui prend toujours soigneusement ses notes se rapprochent, c’est que x considère que les deux situations appartiennent à la même classe d’équivalence.

Il est difficile de ne pas admirer le raffinement méthodologique et conceptuel d’une étude de ce genre : on vérifie la capacité de prédire la variation du niveau de cohérence dans tous les couples formés à partir de six situations relatives au même trait et à partir de quatre stratégies différentes pour évaluer la perceptions idiosyncrasique de chaque individu. Si percevoir les choses de manière personnelle a des conséquences précises sur la cohérence du comportement, il est difficile que cela puisse échapper à des paradigmes expérimentaux de ce genre. La majorité des défenseurs récents de

la théorie de la vertu contre les critiques venant de la psychologie ont accusé la recherche empirique de partir de présupposés simplistes et ne pas prendre en considération des possibles explications alternatives qui sauteraient aux yeux. Par exemple, Weber et Annas accusent les psychologues de présupposer une conception béhavioriste des traits, où avoir un trait est réagir mécaniquement face à un stimulus déclencheur ; Kamtekar se demande pourquoi les psychologues n'ont pas imaginé que, dans le paradigme de Hartshorne et May, les enfants pourraient percevoir différemment une situation où il s'agit de tricher sur une auto-évaluation de vitesse dans un test et une situation où on a l'opportunité de copier de la feuille des réponses³⁵⁶. Maintenant on est en mesure de voir que – au fil des années et du déroulement du débat – la psychologie de la personnalité non seulement n'a pas présupposé une conception unique et simpliste des traits, mais elle a fait son mieux pour explorer avec raffinement méthodologique toute conception possible des traits et plusieurs hypothèses sur la perception et catégorisation des situations. Loin de partir d'une position de supériorité, la philosophie a beaucoup à apprendre de la psychologie de la personnalité sur ces sujets.

Quels ont été les résultats de cette expérience ? Plutôt réconfortants pour l'hypothèse des traits subjectifs. En moyenne, trois des quatre stratégies de catégorisation subjective des situations proposées ont montré une bonne capacité de prédire une cohérence plus importante dans le comportement. La meilleure performance a été celle des méthodes mobilisant le profil idéal, ensuite la classification par rapport à des buts de l'individu et, étrangement, l'évaluation directe de la similitude entre des couples de situations en dernière. Encore plus intéressant que ce résultat moyen, il semble que ces méthodes différentes n'ont pas eu le même succès pour des individus différents, chacune des stratégies ayant donné les meilleurs résultats dans l'un des quatre sous-groupes de taille différente parmi les 40 participants. Cela semble une confirmation du caractère très subjectif de la perception des situations : apparemment des individus différents vont jusqu'à catégoriser les situations sociales avec des principes d'organisation non équivalents, et cela a une conséquence pour la cohérence de leur comportement.

Funder a réalisé une autre expérience qui teste sérieusement l'hypothèse des traits subjectifs. Dans le même paradigme de l'étude citée ci-dessus³⁵⁷ – où des étudiants ont deux conversations libres de

³⁵⁶ Julia Annas, "Virtue ethics and social psychology," *A Priori* 2 (2003), Rachana Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character," *Ethics: An International Journal of Social, Political, and Legal Philosophy* 114, no. 3 (2004), Jonathan Webber, "Virtue, Character and Situation," *Journal of Moral Philosophy* 3, no. 2 (2006).

³⁵⁷ Voir le paragraphe du présent chapitre « Funder : une tentative manquée de dépasser le coefficient de personnalité » et Funder et Colvin, "EXPLORATIONS IN BEHAVIORAL CONSISTENCY - PROPERTIES OF PERSONS, SITUATIONS, AND BEHAVIORS."

cinq minutes (à deux semaines de distance) avec d'autres étudiants qu'ils ne connaissent pas (la deuxième est avec un autre étudiant), et le tout va être filmé et leur comportement évalué sur un large nombre de traits – une évaluation directe de la similitude entre les deux situations de la part des participants a montré une bonne capacité de prédiction du niveau de cohérence dans le comportement effectif évalué. Cette fois la seule stratégie de catégorisation qui a été contrôlée était celle qui n'avait pas bien marché dans l'expérience de Lord (évaluation explicite du niveau de similitude), mais les résultats ont été positifs³⁵⁸.

Faut-il conclure que la voie des traits subjectifs est prometteuse ? Deux raisons barrent la route à un tel optimisme. La première, c'est qu'à notre connaissance il y a eu trop peu d'études encourageantes et méthodologiquement solides. Sauf erreur de notre part, celles que nous avons citées risquent d'être les seules. En plus, la plus élaborée (Lord) n'a testé que le trait du caractère consciencieux, et l'étude de Funder n'a exploré qu'une seule stratégie de catégorisation possible. La deuxième raison est que, pour retenir l'hypothèse des traits subjectifs, il faut davantage qu'une preuve de la capacité prédictive de la perception idiosyncrasique individuelle. Ce dont on a besoin sont des données sur le *poids relatif* de la variation subjective par rapport au pouvoir de prédiction de la similitude objective des situations. Prenons un exemple : pour prouver le caractère véritablement subjectif du trait de l'irascibilité il ne suffit pas de montrer que parfois les gens ont une perception différente de ce qu'est une situation menaçante. Il faut plutôt montrer que l'espace de variation individuelle est très grand par rapport à l'espace d'uniformité (les situations que tout le monde perçoit comme menaçantes). Une variation idiosyncrasique locale et circonscrite dans l'interprétation des situations est certainement possible, réaliste et compatible avec le caractère essentiellement objectif d'un trait. Les traits subjectifs sont ceux où la marge de variation individuelle dans la catégorisation des situations est très large, et le noyau commun restreint. Autrement dit, tous les traits sont certainement en partie subjectifs : le défenseur de la stratégie des traits subjectifs ne peut pas se contenter de confirmer cela par ses expériences. Le fait que les traits aient une petite dimension subjective ne permet pas de montrer que le comportement des gens est cohérent. Or, nous ne disposons pas de données suffisantes pour montrer que la dimension subjective serait plus que minoritaire.

Que les traits qui en seraient résultés se révèlent à l'analyse finalement objectifs est la position de Allport, d'après lequel il y a « *certaines modes de réactions qui sont **approximativement** les mêmes*

³⁵⁸ R. M. Furr et D. C. Funder, "Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches," *Journal of Research in Personality* 38, no. 5 (2004).

d'un individu à l'autre ». Cela semble aussi être bien confirmé par les données. Nous avons vu que dans les expériences de Hartshorne et May, la similarité objective (définie par la présence d'éléments communs dans deux situations : deux tests en classe au lieu d'un test en classe et un à l'extérieur, deux examens au lieu d'un examen et l'occasion de tricher et mentir ailleurs) était bien corrélée avec le niveau de cohérence, même si ce dernier était insatisfaisant aussi dans les cas les plus favorables. Funder a conduit une expérience pour évaluer la capacité prédictive du niveau de similarité objective entre situations par rapport à la cohérence comportementale. Le paradigme était le suivant :

« Un total de 184 étudiants (92 filles, 92 garçons) de l'université de Californie à Riverside ont participé à plusieurs phases de collection de données. En plus, chacun avait recruté un ami(e) du même sexe pour qu'il/elle participe avec lui à des interactions filmées. (...) Chaque participant s'engageait dans six interactions dyadiques, chacune desquelles durait environ cinq minutes et était filmée avec la pleine approbation du participant. Les situations étaient différentes pour deux aspects – partenaire et tâche. La première interaction était une tâche libre dans laquelle les deux participants étaient assis sur un canapé et avaient cinq minutes (...). Cette interaction typiquement se développait dans une conversation pour briser la glace. La deuxième interaction était une tâche coopérative dans laquelle les participants travaillaient ensemble pour construire un modèle, et la troisième interaction était une tâche compétitive dans laquelle le gagnant d'une série de jeux de mémoire recevait un dollar. Les participants ont rencontré ces trois situations (situations 1-3) d'abord avec un inconnu du sexe opposé et après, plusieurs jours plus tard, avec un ami du même sexe qu'ils avaient recruté (situations 4-6)³⁵⁹. »

Les participants étaient après évalués sur une longue liste de traits, comme dans les expériences de Funder décrites ci-dessus. On s'intéressait à voir si la variation des deux paramètres qui distinguaient les situations (partenaire et tâche) était en mesure de prédire les variations de cohérence. Les résultats ont été plutôt positifs, montrant que la cohérence est inversement proportionnelle à la différence objective entre des situations.

Cette étude, tout comme les classiques dans le débat sur la cohérence inter-situationnelle, conceptualise le niveau de similitude objectif entre des situations à partir de facteurs objectifs qu'elles partagent : même lieu, partenaire, tâche. Cela représente la stratégie standard dans la littérature : par exemple, Magnusson, Gerzen et Nyman ont mesuré la cohérence en faisant varier la

³⁵⁹ Ibid.: 434.

composition du groupe et la nature de la tâche³⁶⁰. Toutefois, nous avons vu dans la section précédente, où il s'agissait de faire de découvrir la façon dont les gens catégorisent les situations (dans l'espoir que cela aide à expliquer le bas niveau de cohérence dans les expériences), qu'il y a plusieurs façons de conceptualiser la perception des situations des gens. Au lieu de chercher des surprises par rapport à nos intuitions théoriques sur ce qui explique la similitude des situations, on peut utiliser ces différentes conceptualisations pour tester l'association entre similitude et cohérence.

Dans la littérature ont été explorées deux stratégies différentes. La première consiste à définir la similarité objective entre deux situations comme la moyenne dans un groupe de l'évaluation subjective de similarité par chacun des membres, quelle que soit la méthode choisie pour établir cette dernière. Ainsi, Klirs et Revelle ont trouvé que la moyenne d'évaluations subjectives directes est en mesure de prédire le niveau de cohérence³⁶¹. Ce résultat n'a pas été répliqué par Lord. Dans la même expérience qu'on a décrite ci-dessus, le pouvoir de prédiction des valeurs moyennes à partir de l'ensemble des quatre méthodes d'évaluation subjective (similarité explicite, classement par buts de l'agent, différence entre le profil de l'agent et le profil idéal de la situation, différence entre les profils idéaux des situations) a été testé³⁶². Sur la base de chacun de ces méthodes, il s'agit simplement de faire une moyenne des estimations subjectives de similarité. Ces valeurs moyennes n'ont pas été en mesure de prédire les variations de cohérence. Dans l'expérience de Klirs il s'agissait uniquement de confronter la catégorisation de situations et des *déclarations* sur ce qu'on aurait fait dans la situation donnée (ce qui n'est pas fiable) : le bon résultat compte moins que le mauvais de Lord, obtenu avec des vraies données comportementales. Nous devons conclure que cette stratégie s'est avérée un échec.

Une deuxième stratégie a été explorée par Shoda et Mischel, dans une des analyses à partir de la grande base de données du camp estival de Wediko, l'étude la plus sophistiquée qui ait jamais été faite sur la structure et cohérence de la personnalité, et dont nous aurons l'occasion de parler à plusieurs reprises³⁶³. Il s'agit d'une expérience où l'activité d'un groupe d'enfants et pré-

³⁶⁰ D. Magnusson, M. Gerzen et B. Nyman, "GENERALITY OF BEHAVIORAL DATA-I - GENERALIZATION FROM OBSERVATIONS ON ONE OCCASION," *Multivariate Behavioral Research* 3, no. 3 (1968).

³⁶¹ Klirs et Revelle, "PREDICTING VARIABILITY FROM PERCEIVED SITUATIONAL SIMILARITY."

³⁶² Lord, "PREDICTING BEHAVIORAL CONSISTENCY FROM AN INDIVIDUALS PERCEPTION OF SITUATIONAL SIMILARITIES."

³⁶³ Y. Shoda, W. Mischel et J. C. Wright, "THE ROLE OF SITUATIONAL DEMANDS AND COGNITIVE COMPETENCES IN BEHAVIOR ORGANIZATION AND PERSONALITY COHERENCE," *Journal of Personality and Social Psychology* 65, no. 5 (1993).

adolescents avec problèmes comportementaux liés à l'agressivité et au manque de maîtrise de soi a été presque entièrement filmée, analysée et codée pendant les six semaines d'un camp estival. Les expérimentateurs ont développé l'idée qu'une situation sera perçue de manière différente selon le genre de capacité nécessaire pour l'affronter : cela rend possible une classification des compétences qui sont activées par différentes situations. Dans la taxinomie qu'ils offrent (adaptée aux activités des enfants dans le camp), il y a cinq catégories de compétence : cognitives (des compétences telles que : « penser rationnellement, logiquement », « se concentrer », « inférer des relations de cause et effet », « montrer un niveau élevé d'intelligence »), sociales (« parler en public », « exprimer ses propres émotions », « résoudre des conflits avec d'autres enfants », « résoudre des disputes avec des adultes »), maîtrise de soi (« supporter un délai de gratification », « tolérer la frustration »), force physique (« force physique », « résistance aux blessures », « vitesse, agilité »), coordination (« coordonner le mouvement du corps avec le bras », « adresse »). L'étude s'intéresse à deux traits presque opposés, l'agressivité et l'attitude pro-sociale (ex : « commencer une conversation », « chercher des contacts »). Le niveau moyen d'agressivité et de sociabilité a été évalué pour chaque enfant (60 garçon et 24 filles, âgés de 6 à 12, moyenne 10 ans et 2 mois) dans chacune de 14 activités différentes, dont par exemple : activité artistique, athlétique, réunion, natation, interaction avec la nature, jeux de groupe etc. La moyenne pour chaque activité et enfant se base sur une observation de l'enfant pendant un certain nombre d'heures, où les épisodes pertinents étaient isolés et évalués, toujours avec un bon niveau d'accord entre plusieurs évaluateurs. Les 14 activités étaient aussi évaluées de façon consensuelle sur une échelle 0-7 par rapport aux compétences requises, suivant la taxinomie présentée. Cela a permis de tester la capacité de prédiction de la similarité entre situations (ainsi définie) par rapport au niveau de cohérence manifesté dans les 14 activités différentes. Surtout pour le trait d'agressivité, les résultats ont été positifs, ce qui est remarquable étant donné le nombre considérable de situations testées. Avec un nombre limité de situations, il est possible qu'une preuve de cohérence se révèle comme un artefact expérimental : deux situations perçues différemment pourraient par hasard être associées aux mêmes réactions dans la structure de personnalité du sujet, dans ce cas, une attitude idiosyncrasique (je sauve une vieille dame car elle ressemble à ma grand-mère, je sauve un enfant parce que je le trouvais sympathique) serait prise pour du courage depuis l'extérieur. Mais avec 14 situations différentes, ce genre de problème ne peut pas se présenter.

Moralité ? La perception des situations semble être essentiellement homogène parmi des individus différents, la variation subjective existe mais paraît plutôt locale : des gens irascibles trouvent menaçantes des situations différentes, mais il y a certainement un gros noyau de situations où la

menace est perçue par tout le monde. Comme le dit Funder : « *Au niveau théorique, même si l'extension des perceptions potentiellement idiosyncrasiques est illimitée en principe, en pratique il semble probable que la plupart des gens vont interpréter certaines situations de la même façon*³⁶⁴. » En conclusion, même si la question reste ouverte à des nouvelles études empiriques (comme c'est toujours le cas, mais spécialement en psychologie de la personnalité), il ne semble pas crédible d'expliquer les bas coefficients de corrélations pour la cohérence inter-situationnelle par le fait que les traits seraient subjectifs (ou beaucoup plus subjectifs qu'on ne le croit). Les études empiriques très sophistiquées qu'on vient d'analyser n'étaient pas l'idée qu'on réagirait tous de la même façon de manière cohérente, *si ce n'était qu'on perçoit les situations différemment*. Les incohérences dans nos actions paraissent plus profondes.

Méta-Traits

Dans la quatrième hypothèse, des signes de désistement commencent à se manifester dans l'attitude des psychologues de la personnalité. En effet, il s'agit d'admettre que le niveau élevé de cohérence dont la recherche a marqué l'histoire de la discipline se manifeste dans certaines personnes plutôt que d'autres, et dans certains traits de caractère plutôt que d'autres : pour le reste, on pourrait être contraint de concéder que le comportement des gens n'est pas très cohérent. Il ne faut pas confondre l'hypothèse avec la platitude que certaines personnes sont (par exemple) courageuses et d'autres non : cela est accepté par définition par tout psychologue de la personnalité ! Dans l'approche traditionnelle, une personne qui n'est pas courageuse est conçue comme étant constamment frileuse ou constamment « moyennement courageuse », alors qu'ici il s'agit de dire que dans certaines dimensions de leur personnalité (la gestion du danger) les gens ne manifesteraient aucune structure cohérente, du tout. On pourrait être très peu courageux mais très cohérent dans ce manque de courage, alors qu'un individu qui ressort comme moyennement courageux dans un questionnaire de personnalité pourrait ne manifester aucune cohérence. Ce genre d'hypothèse dans la littérature psychologique est connue sous les noms de « approche de la variable modératrice » (*moderator variable approach* : car il y aurait des variables qui expliquent pourquoi certains traits sont associés à certaines personnes), « approche des méta-traits » (car si tout le monde ne se place pas de manière cohérente à un certain niveau dans l'échelle de tous les traits, le fait de pouvoir être véritablement décrit par des traits devient un trait en soi-même, le méta-trait d'avoir des traits) ou, de façon plus

³⁶⁴ Furr et Funder, "Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches," 424.

informelle « certaines personnes quelquefois » (*some people some of the time*)³⁶⁵. Il s'agit d'une hypothèse très importante pour le lien entre psychologie de la personnalité et théorie de la vertu, car cette dernière n'a généralement pas supposé que le comportement de tout le monde est cohérent : la cohérence est justement le signe distinctif du vertueux.

« ...les théoriciens traditionnels de la vertu se plaignent fréquemment du fait que la plupart des gens sont instable et incohérents. Platon, par exemple, dit que les gens non vertueux ne sont "jamais les mêmes, pas même par rapport à soi. Ils sont impulsifs et instables" ; une personne qui n'a pas la vertu de la justice "n'a pas un seul esprit" mais "se balance en avant et en arrière". De façon similaire, Epictète réprimande son auditoire, "vous êtes fréquemment abasourdis et perturbés par vos impressions extérieures, et dépassés par leur nature persuasive ; et à un moment vous considérez ces choses comme bonnes, et après à nouveau vous les considérez comme mauvaises – même si elles sont exactement les mêmes – et après encore comme ni bonnes ni mauvaises ; et, en un mot, vous êtes vulnérables à la souffrance, à la peur, à l'envie, à l'agitation et au changement"³⁶⁶. »

L'approche des méta-traits joue exactement dans le même camp que la théorie de la vertu.

Pour preuve du décalage entre l'arrière-plan théorique et la pratique expérimentale qui se produit souvent, l'hypothèse des méta-traits était déjà présente chez le père de la psychologie de la personnalité, Gordon Allport.³⁶⁷ Dans son approche idiographique, il avait soutenu que les traits les plus centraux dans la conduite et dans l'auto-perception d'une personne seraient associés à un niveau de cohérence plus élevé. Comme le résume très bien Mischel :

« Allport pensait que certaines personnes ont des dispositions qui influencent la plupart des aspects de leur comportement. Il appelait ces dispositions très générales traits cardinaux. Par exemple, si toute la vie d'une personne semble s'organiser autour de la poursuite de ses buts et de l'excellence, alors la détermination à atteindre ses objectifs pourrait être son trait cardinal. Les traits centraux sont des dispositions moins omniprésentes mais encore plutôt générales, et Allport pensait que beaucoup de gens sont largement influencés par des traits

³⁶⁵ Bem et Allen, "ON PREDICTING SOME OF PEOPLE SOME OF TIME - SEARCH FOR CROSS-SITUATIONAL CONSISTENCIES IN BEHAVIOR."

³⁶⁶ Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character," 482.

³⁶⁷ Allport, *Personality; a psychological interpretation*. J'insiste sur le fait que la littérature expérimentale est conceptuellement et méthodologiquement raffinée *dans sa globalité*. C'est toujours à ce niveau qu'il faut évaluer des traditions expérimentales. A un moment donné il peut toujours y avoir un décalage entre des hypothèses intéressantes et une pratique expérimentale qui n'a pas les moyens suffisants pour les tester.

centraux. Finalement, des traits plus spécifiques et limités sont appelés dispositions secondaires ou "attitudes"³⁶⁸. »

L'hypothèse est restée dormante jusqu'à la critique dévastatrice de l'approche des traits par Mischel, qui a motivé Bem et Allen à explorer cette voie³⁶⁹. Une théorie des méta-traits doit faire deux choix fondamentaux : la dimension théorique qui expliquerait le fait d'être ou de ne pas être évaluable dans l'échelle d'un trait (la propriété psychologique qui explique le méta-trait), et une définition opérationnelle de cette dimension. Dans la littérature, plusieurs solutions ont été proposées pour les deux niveaux, mais il faut remarquer que dans sa globalité cette tradition de recherche est restée très fidèle à Allport.

Dans l'absolu, il est possible de concevoir une longue liste de propriétés psychologiques hétérogènes qui pourraient expliquer le méta-trait d'avoir un trait. On pourrait imaginer que chacun développe les traits qui sont les plus utiles dans son environnement, ou certains plutôt que d'autres à cause de l'influence des parents ou d'événements traumatiques etc. Il n'est pas évident que le rôle central dans la structuration de la conduite et de l'auto-perception soit le facteur central. Or, la proposition d'Allport restant floue et sous-déterminée, les études contemporaines ont surtout exploré une série d'approximations non équivalentes à l'idée d'enracinement et de centralité d'un trait dans l'auto-perception et le schéma du « soi », alors que l'éventail des possibilités aurait pu être plus large. On voit qu'une approche à la Allport sur la nature des méta-traits ne fait qu'approfondir le rapprochement de la psychologie de la personnalité avec la théorie des vertus. La vertu est censée être rare, permettre un niveau de cohérence élevé par rapport aux incertitudes et hésitations des gens communs et aussi réorganiser complètement la personnalité du vertueux. La vertu (ou les différentes vertus acquises) devient le cœur de la personnalité du vertueux, et s'y manifeste massivement. Même s'il y a certainement beaucoup à ajouter, les vertus remplissent les critères des traits cardinaux de Allport. La recherche par des moyens expérimentaux de traits cardinaux très cohérents devient l'un des meilleurs points de contact direct entre théorie de la vertu et psychologie empirique.

Parmi les différentes notions qui ont été proposées pour articuler l'intuition de Allport, on peut compter « l'importance et la pertinence subjectives », la « cohérence perçue » et la « perception de l'unicité d'un trait donné » (ex : peu de gens sont consciencieux comme je le suis). D'autres notions proposées mais qui s'écartent davantage de l'approche de Allport sont le « niveau de monitoring de

³⁶⁸ Mischel, *Introduction to personality : toward an integration* 49.

³⁶⁹ Bem et Allen, "ON PREDICTING SOME OF PEOPLE SOME OF TIME - SEARCH FOR CROSS-SITUATIONAL CONSISTENCIES IN BEHAVIOR."

soi » (qu'on va analyser ci-dessous) et le « niveau d'observabilité d'un trait donné ». En ce qui concerne les définitions opérationnelles de ces notions, il y a eu également une variété de méthodes : les plus utilisées ont été l'auto évaluation directe sur une échelle numérique (« dans quelle mesure le trait x est-il central dans ta personnalité ? », « êtes-vous très cohérent dans le comportement associé au trait x ? », « dans quel mesure êtes-vous unique en ce qui concerne ce trait ? ») ou un classement d'une liste de traits en ordre d'importance pour le sujet (pour mesurer l'importance et la pertinence subjectives), pour le niveau de monitoring de soi, un questionnaire spécifique a été utilisé. Une autre définition opérationnelle du niveau du méta-trait correspondant à un trait donné se base sur la variation inter-item dans un questionnaire, mesuré par la déviation standard dans la réponse aux différentes questions qui composent les sections du questionnaire. Il s'agit d'une méthode qui a eu beaucoup de succès car elle est très pratique, à partir des mêmes réponses qui permettent d'obtenir le niveau dans un trait (et la littérature sur les questionnaires est énorme), on peut calculer la déviation standard, sans manipulations expérimentales supplémentaires. Ce n'est pas une mesure qui est conceptuellement liée en particulier à une des dimensions psychologiques qu'on vient d'introduire, au contraire, exprimer une grande variété dans les réponses (censées être liées au même trait) paraît un symptôme général du manque de méta-trait. A son tour, cette mesure permet de vérifier la validité des définitions de départ : si la « cohérence perçue » n'est pas associée de façon intelligible à la variabilité inter-item, c'est que ce n'est pas une bonne explications des méta-trait.

Il ne faut pas voir cette prolifération des dimensions censées expliquer les méta-trait et de définitions opérationnelles différentes comme un signe de confusion méthodologique. On pourrait penser que c'est le cas, vu que tantôt on part d'une hypothèse explicative (ex : méta-trait = pertinence subjective) et on cherche une mesure de cette dimension (ex : auto-évaluation de la pertinence), tantôt on part d'une définition opérationnelle (ex : méta-trait = variation inter-item) et on cherche à établir si certaines dimensions psychologiques s'y conforment (ex : on se demande si un bas niveau de pertinence subjective est associé à une grande variation inter-item). Il s'agit au contraire d'une stratégie « cohérentiste » qui caractérise profondément la psychologie de la personnalité. Pour prouver l'existence et l'influence de facteurs difficiles à caractériser de manière précise, comme le niveau d'intégration de la représentation d'un trait dans la représentation qu'on a de soi-même, on cherche à créer un réseau d'interactions entre des caractérisations différentes et parfois intentionnellement indirectes de la dimension psychologique concernée et des mesures intentionnellement non équivalentes de ses effets. C'est l'interprétation finale de ce réseau de donnée accumulées par la recherche qui permet de prouver l'influence étudiée : si plusieurs facteurs non équivalents mais qui relèvent de la même dimension psychologique plus générale sont associés

à une classe d'effets non identiques mais liés, alors que des facteurs plus éloignés de la dimension psychologique de départ produisent des résultats significativement différents, on a une preuve de l'effet général cherché.

La même approche cohérentiste est à la base de l'élaboration de tous les questionnaires censés mesurer le niveau d'un certain trait : supposons que l'on veuille mesurer « l'attrance pour les expériences extrêmes ». Les psychologues ne donnent pour acquis ni que le trait existe (tout trait arbitraire qu'on peut définir n'existe pas : un trait existe si les gens sont positionnés de façon relativement stable à un certain niveau de son échelle) ni qu'on sait comment le mesurer s'il existe. Comment sortir de l'impasse ? On procède de façon cohérentiste : on part d'une hypothèse sur la façon de mesurer le trait (on compile un questionnaire) et on voit si cela mesure quelque chose de stable (fiabilité test-test), après on étudie les propriétés de ce trait *x*. S'il se comporte comme s'il s'agissait de « l'attraction pour les expériences extrêmes » (il est associé de façon intelligible à d'autres traits qui paraissent proches) on peut en conclure en même temps que le trait existe et qu'on sait comment le mesurer, et on dit que la version finale du questionnaire a été validée. Si ce qui est mesuré par le questionnaire se comporte d'une façon qui n'a rien à voir avec les expériences extrêmes, alors il faut changer le questionnaire et recommencer. Il faut aussi que le trait se comporte de façon suffisamment différente par rapport à d'autres traits déjà connus, pour éviter de l'identifier avec ces traits (*discriminant validity*). S'il se trouve que les gens qui sont attirés par les expériences extrêmes sont tout simplement des extraverties qui ont beaucoup d'argent, on n'aura pas le droit de postuler l'existence du nouveau trait « attrance pour les expériences extrêmes ». A la limite, si le processus de validation n'aboutit pas à des résultats, on peut en conclure que le trait n'existe pas³⁷⁰. En ce qui concerne la notion de méta-trait, quel a été le résultat du processus de validation ? Dans la littérature, cette panoplie de définitions opérationnelles ont été associées à trois variables « critère » censées mesurer le niveau de cohérence effective du comportement : le niveau d'accord entre plusieurs évaluateurs du sujet (accord inter-juges), la corrélation entre l'évaluation et des manifestations comportementales (trait-comportement) et la corrélation entre plusieurs

³⁷⁰ Il y a très peu de notions psychologiques qui sont bien validées. Parmi les notions qui sont accusées d'avoir une validité modeste, on en trouve de célèbres. Par exemple, dans le débat sur l'intelligence qui oppose les défenseurs de l'intelligence générale aux défenseurs de formes spécifiques d'intelligence (les huit types d'intelligence proposés par Gardner ou les trois types proposés par Sternberg – voir le chapitre sur la sagesse, paragraphe consacré à Sternberg), il faut que reconnaisse que, d'un point de vue strictement technique, la « validité » des formes spécialisées d'intelligence est plutôt faible. En pratique, cela signifie qu'on n'arrive pas à le mesurer de façon fiable, et à formuler des prédictions informatives. Sur cette question voir Geher et Miller, *Mating Intelligence: Sex, Relationships, and the Mind's Reproductive System* 4.

manifestations comportementales (comportement-comportement). Globalement, les résultats sont intéressants dans l'absolu mais décevants pour les ambitions de la psychologie de la personnalité et de la théorie de la vertu. En bref, les méta-traits existent probablement, mais au lieu de rendre possible le niveau élevé de cohérence recherché, ils se révèlent nécessaires pour atteindre les corrélations modestes déjà présentes dans la littérature : les gens qui ont un méta-trait du courage très prononcé pourront atteindre des corrélations inter-situationnelles de 0.3 ou 0.4 pour le trait du courage, corrélations qui tombent pour les autres.

Le plus gros problème de cette littérature est que les meilleurs résultats et les efforts les plus récents (environ tout ce qui a été fait après les années 80) sont associés à l'étude de la corrélation inter-juges. Or, nous savons déjà que l'accord sur l'évaluation d'un sujet cible de la part de personnes qui le connaissent bien (mais qui ne se connaissent pas forcément entre elles) peut être remarquablement élevé. Mais cela ne nous informe pas sur le niveau de cohérence effectif du comportement de la personne : le consensus prouve uniquement la présence des mêmes processus de catégorisation qui répondent aux mêmes signes comportementaux. Si, comme ces études le montrent, il y a plus de consensus sur comment évaluer des personnes qui se considèrent uniques par rapport au trait concerné, ou qui le voient comme central dans la perception de soi, ou encore qui ont la perception subjective d'être cohérents, cela signifie que ces perceptions sont associées à des « signes » réels dans le comportement du sujet³⁷¹. Ces signes sont perçus par les observateurs qui apparemment les catégorisent de la même manière, ce qui justifie le niveau de consensus plus fort parmi les évaluateurs. Mais rien ne nous autorise à inférer que le signe perçu soit la cohérence inter-situationnelle effective. Argumenter que la meilleure explication du consensus est la vraie existence du trait n'est qu'une pétition de principe dans une analyse détaillée de la cohérence du comportement. Nous savons depuis le début du débat sur la cohérence que le consensus parmi les évaluateurs peut être remarquable³⁷².

En réalité, comme nous le verrons plus tard, l'argument qui relie le consensus et la cohérence peut être développé de façon plus subtile. L'un des acquis les plus intéressants des dernières décennies de la psychologie de la personnalité est un corpus très riche de données sur les facteurs qui favorisent ou limitent le consensus de plusieurs évaluateurs sur un sujet cible³⁷³. Or, si l'accord en soi ne nous informe pas sur les signes comportementaux qui attirent l'attention des gens, toute

³⁷¹ D. T. Kenrick et D. C. Funder, "PROFITING FROM CONTROVERSY - LESSONS FROM THE PERSON-SITUATION DEBATE," *American Psychologist* 43, no. 1 (1988).

³⁷² Mischel, *Personality and assessment* 47.

³⁷³ D. C. Funder, "ON THE ACCURACY OF PERSONALITY JUDGMENT - A REALISTIC APPROACH," *Psychological Review* 102, no. 4 (1995).

hypothèse sur la nature des structures psychologiques sur lesquelles porte l'accord n'est pas compatible avec la connaissance détaillée de facteurs qui influencent le niveau de consensus. Par exemple, si des gens que connaissent très bien le sujet sont sensiblement plus en accord que de simples connaissances, et si l'accord est plus élevé pour des traits qui sont plus facilement observables, il paraît plus plausible que les signes qui stimulent l'accord soient effectivement liés à des actions pertinentes pour le trait. En d'autres termes, si l'accord était expliqué par le fait que tout le monde stéréotype les autres de la même façon, on ne voit pas pourquoi les proches devraient avoir un avantage sur les simples connaissances. L'accord se comporte « comme si » il suivait les bons indices.

Cet argument est ingénieux mais il montre beaucoup moins que l'on ne croit. Ce qui est en effet probable est que le jugement sur le niveau de générosité se base sur des actions qui relèvent vraiment du trait, c'est-à-dire des actions généreuses. Ce qui exclut l'hypothèse qu'on stéréotype de façon complètement infondée à partir de signes qui ne sont pas liés à la notion de générosité. Mais il se peut très bien que ce sur quoi l'accord porte soit le fait d'être plutôt généreux en moyenne, en général, comme tendance. Il est tout à fait plausible qu'il faille connaître bien une personne pour être sûrs de ses tendances moyennes dans les aspects le plus manifestes de son comportement. Rien ne départage cette hypothèse de l'idée de traits parfaitement cohérents, et comme notre but est celui d'évaluer la possibilité d'une cohérence forte, la différence est fondamentale. En conclusion, même après tout ce qu'on a découvert d'extrêmement intéressant sur la nature du consensus entre juges différents, la question reste non pertinente pour résoudre le problème de la cohérence inter-situationnelle.

Des données directes sur le lien entre méta-traits et comportement existent (comportement-comportement), et elles laissent très peu d'espoir : l'effet est petit et souvent n'a pas été répliqué, en tout cas beaucoup moins que pour les études sur l'accord entre juges. Dans l'article de Bem et Allen, on ne trouve pas les valeurs spécifiques des corrélations comportement-comportement et trait-comportement. Elles sont confondues dans une moyenne générale des mesures de cohérence qui prend en compte aussi l'accord entre juges. En tout cas, les mesures concernaient uniquement deux traits : le caractère consciencieux et le caractère amical. Mischel et Peake n'ont pas répliqué le résultat dans une étude où la cohérence inter-situationnelle était calculée pour le caractère consciencieux à partir de six situations différentes de la vie d'un groupe d'étudiants dans une université américaine : l'effet était visible au niveau d'accord inter-juges (corrélations 0,68/0,22 respectivement dans le sous groupe caractérisé par le méta-trait et dans le reste) mais absent au

niveau de cohérence inter-situationnelle directe (0,11/0,14)³⁷⁴. Dans une étude où la cohérence inter-situationnelle a été mesurée pour huit traits, le résultat a été également décevant, avec des valeurs très peu élevées et incohérentes, par exemple : amicalité (0,3/-0,3), caractère consciencieux (-0,06, 0,07), stabilité émotionnelle (0,13/0,11)³⁷⁵. Ces valeurs concernent les corrélations avec des méta-traits mesurés par auto-évaluation directe (ex « quel est ton niveau de cohérence dans ce trait ?), mais dans les mêmes deux articles des résultats de la même teneur sont reportés pour d'autres méthodes de mesure des méta-traits.

En ce qui concerne la corrélation trait-comportement, la situation est un peu plus positive (cela ne résout pas le problème posé par les mauvaises données comportement-comportement, seule preuve directe de cohérence inter-situationnelle). L'existence de l'effet a été répliquée de façon plus régulière (mais, par exemple, pas dans l'article de Chaplin sur huit traits de caractères qu'on vient de citer, où les corrélations sont de la même teneur que celle comportement-comportement) : un niveau élevé du méta-trait d'une variante d'autocontrôle (*locus of control*) a augmenté la capacité de prédiction de la persistance dans une tâche de laboratoire, et le méta-trait de l'estime de soi a été directement associé à la disposition à demander conseil (*advice seeking*)³⁷⁶. Dans le même registre, même si cela concerne davantage la corrélation avec des résultats généraux et non avec des comportements précis détectables (ce qui introduit un niveau de médiation ultérieur par rapport au problème de la cohérence inter-situationnelle), on peut citer la capacité des méta-traits d'améliorer la capacité de prédiction de résultats objectifs au travail, comme cela a été démontré dans une étude sur la performance d'employés dans une entreprise électrique³⁷⁷. L'effet ici est très intéressant, avec une différence 0,37/0,07 dans les corrélations entre le sous-groupe qui possède le méta-trait du « suivre les procédures de sécurité de façon consciencieuse » et les autres. Néanmoins, Il faut remarquer que 0,37 reste très proche et n'est pas une violation flagrante du célèbre plafond de 0,30 postulé par Mischel. Plus important, on parle ici d'un trait plutôt restreint et « scripté³⁷⁸ », ce n'est

³⁷⁴ Mischel et Peake, "Beyond Deja-Vu in the Search for Cross-Situational Consistency," 747.

³⁷⁵ W. F. Chaplin et L. R. Goldberg, "A FAILURE TO REPLICATE THE BEM AND ALLEN STUDY OF INDIVIDUAL-DIFFERENCES IN CROSS-SITUATIONAL CONSISTENCY," *Journal of Personality and Social Psychology* 47, no. 5 (1984): 1084.

³⁷⁶ R. F. Baumeister et D. M. Tice, "METATRAITS," *Journal of Personality* 56, no. 3 (1988).

³⁷⁷ S. A. Dwight, P. P. Wolf et J. H. Golden, "Metatraits: Enhancing criterion-related validity through the assessment of traitedness," *Journal of Applied Social Psychology* 32, no. 10 (2002).

³⁷⁸ Un script est un plan, une procédure ou un schéma d'action qui définit les étapes principales d'une tâche ou d'une interaction sociale, sans établir nécessairement tous les détails. Par exemple, une interaction avec un serveur dans un restaurant est « scriptée ». On commence par demander le menu, après on commande, après on demande l'addiction. La plupart des questions qu'on peut poser à un serveur sont typiques (« où peut-on acheter des

pas le caractère consciencieux en général, mais « suivre des *procédures* de sécurité de façon consciencieuse ». La question qui nous concerne dans ce chapitre est la possibilité d'un niveau de cohérence élevé de trait « larges », non pas la cohérence du « caractère consciencieux au travail » qui relève déjà d'un renoncement par rapport à la théorie des traits traditionnelle. En plus, il est clair que suivre des procédures relève d'un contexte d'action très « scripté » où un niveau élevé de cohérence est plus facile à atteindre, il ne s'agit pas des habitudes générales et intelligentes que sont les vertus et les traits de caractère.

Une stratégie non néo-Allportienne déployée dans le programme de recherche sur les méta-traités mérite d'être mentionnée, car elle marque un autre rapprochement méthodologique entre psychologie de la personnalité et théorie de la vertu. D'après la plupart des approches philosophiques, non seulement il n'y qu'une minorité de vertueux dans la population, mais la capacité à comprendre le caractère autrui n'est pas également distribuée. Il s'agit d'un thème important dans l'épistémologie de la théorie de la vertu : aucune élaboration théorique de la notion de vertu ne se présente en tant que simple reprise du sens commun, il s'agit plutôt de puiser du fond de sagesse représenté par les usages les plus intelligents du terme et des explications psychologiques relatives à la vertu, de la part de personnes qui ont une grande compréhension pratique du comportement des gens. Le concept de vertu est possédé et compris par tout le monde mais ne peut être maîtrisé que par une minorité de gens. Il y aurait là une forme de divisions du travail épistémique et sémantique typique de notions répandues mais difficiles – comme le concept scientifique d'atome – en ce cas justifié non par la provenance d'un domaine d'expertise spécifique mais par la difficulté d'application d'une notion commune qui requiert une compréhension profonde de l'esprit humain : cela expliquerait la possibilité que l'usage commun des termes de traits soit souvent associé à des stéréotypes.

Dans le cadre du paradigme expérimental qui cherche la corrélation trait-comportement (car la corrélation comportement-comportement – pourvu qu'on ait bien choisi des situations cruciales à tester à partir de la définition du trait concerné – produit des données directes sur la cohérence, indépendamment de la perception des traits), Fazio, Zanna et Olson ont développé cette idée. Ils sont partis du constat que seuls une minorité des gens sont très cohérents et que seuls une minorité

cigarettes ? », « y a-t-il des pharmacies dans le coin ? »), et il a déjà un modèle de réponse préparé. Un trait scripté est un trait de caractère défini par un ensemble de procédures : par exemple, la gentillesse d'une infirmière est scriptée, car l'infirmière va réagir d'une façon plutôt prévisible et structurée dans chacune d'une série de situations typiques qui se présentent dans l'interaction avec les patients.

(pas forcément la même) a une perception adéquate de son propre caractère³⁷⁹. Leur hypothèse de travail pour expliquer le phénomène est que la perception de la personnalité ne vise pas forcément à une précision descriptive : se faire une idée sur soi-même et sur autrui en termes de traits de caractère relève avant tout d'une gestion de l'image dans le groupe, processus qui doit satisfaire à des exigences pragmatiques très fortes et qui reste ouvert à une série de biais perceptifs. Une compréhension de la personnalité fidèle à la réalité est plutôt l'exception que la règle.

La deuxième hypothèse est qu'une auto-évaluation fidèle est possible seulement lorsque deux conditions sont remplies : a. le sujet doit être vraiment cohérent dans les faits et b. il doit baser son jugement sur les faits – en faisant une espèce d'induction portant sur soi-même – et ne pas chercher une gestion de l'image. La définition opérationnelle de cette deuxième attitude a été celle de « niveau de monitoring de soi », mesuré par une échelle spécifique³⁸⁰. L'idée est que les gens vont se faire une idée adéquate d'eux-mêmes s'ils considèrent ce qu'il leur arrive de faire comme une manifestation de leur personnalité (bas niveau de monitoring de soi), alors qu'ils seront égarés s'ils se réservent de penser qu'ils sont toujours différents et qu'ils ont un potentiel inexprimé dans leur actions (niveau élevé). Dans cette conception si une personne fait très attention à ses actions (et garde un niveau d'auto monitoring élevé) c'est pour les interpréter et s'en distancier, alors qu'une attitude spontanée permettrait de bonnes statistiques sur soi-même. En ce qui concerne la première contrainte, on suppose que des gens incohérents abandonneraient en tout cas une attitude réaliste vers eux-mêmes (au lieu de se percevoir fidèlement comme incohérents) à cause de la difficulté de suivre et d'enregistrer son propre comportement instable (qui n'est d'ailleurs pas flatteur).

Sur la base de ces prémisses, les auteurs ont organisé une expérience (n= 103) en deux phases, centrée sur le trait de la « religiosité ». Dans un premiers temps, les participants ont rempli le questionnaire qui mesure le monitoring de soi et ont donné une auto-évaluation sur une échelle 0-6 de leur cohérence dans la dimension de la religiosité. Un mois plus tard, un questionnaire posait une série de (environ 90) questions non pas sur leur attitude générale, mais sur ce qu'ils avaient fait concrètement entretemps pour vivre leur religiosité, avec le but de mesurer la cohérence inter-situationnelle. D'autres questions du même genre, portant sur la prière, le sexe avant le mariage, l'utilisation de drogue et d'alcool dans la période concernée étaient posées séparément. Les résultats ont été très encourageants, avec des corrélations plus élevées d'entre 0,20 et 0,30 par rapport au reste pour le sous-groupe des cohérents et bas niveau de monitoring, comme prévu par l'hypothèse.

³⁷⁹ M. P. Zanna, J. M. Olson et R. H. Fazio, "ATTITUDE-BEHAVIOR CONSISTENCY - AN INDIVIDUAL DIFFERENCE PERSPECTIVE," *Journal of Personality and Social Psychology* 38, no. 3 (1980).

³⁸⁰ M. Snyder, "SELF-MONITORING OF EXPRESSIVE BEHAVIOR," *Journal of Personality and Social Psychology* 30, no. 4 (1974).

Encore plus impressionnantes étaient les corrélations entre les mesures de la première séance et le questionnaire de la deuxième, avec des corrélations de 0,75/0,50 pour les deux sous-groupes. Un espoir d'avoir trouvé la super-cohérence ? Malheureusement, l'explication de l'anomalie est vite trouvée. A la différence des corrélations calculées pour les questions posées séparément (qui étaient sur un niveau classique de 0,30), du questionnaire on avait tiré un score global, il s'agit donc d'une corrélation avec une valeur moyenne, non pas une vraie corrélations inter-situationnelle. A part cela, nous pouvons remarquer que

1. nous ne voyons pas comment dans la première séance le niveau perçu de cohérence dans l'attitude religieuse peut être considéré comme une mesure de cohérence effective, ce qui est requis par l'hypothèse testée
2. la procédure présuppose la sincérité des participants sur ce qu'ils ont fait dans le mois entre les deux séances, et n'a donc pas la même valeur qu'une mesure directe
3. la définition opérationnelle utilisée dans les questions est liée aux actions typiques d'un pratiquant, il s'agit d'un système d'habitudes, de pratiques plutôt codifiées (« scriptées ») où il est plus facile d'être cohérent.

En conclusion, l'analyse d'une série de stratégies développées par les psychologues nous apprend que les méta-traits existent et qu'ils ont la capacité d'augmenter le niveau de cohérence (un effet tantôt faible tantôt modeste)³⁸¹, mais il s'agit plutôt d'atteindre le déjà célèbre plafond d'environ 0,30 que de le dépasser. En ce qui concerne la cohérence inter-situationnelle, cette stratégie se révèle finalement un échec. Cela est très significatif pour la théorie de la vertu : les vertueux ou au moins les gens qui s'approchent de façon significative de la vertu sont censés être un sous-groupe de la population caractérisé par des propriétés qui ne s'éloignent pas trop de celles que les psychologues ont explorées en étudiant les méta-traits. Il est très plausible d'après les caractérisations de la vertu les plus répandues dans la littérature que le vertueux ait une auto-perception adéquate de sa propre cohérence et en général de son caractère, qu'il voit sa vertu comme centrale dans sa vie et dans sa personnalité, qui sont ainsi réorganisées de façon unique.

³⁸¹ Même s'ils ne sont pas associés à la vraie cohérence, les méta-traits – surtout mesurés par la variabilité inter-item – ont des propriétés psychométriques intéressantes. La manifestation d'un certain niveau de variabilité dans les réponses aux questionnaires relatifs à un trait donné est elle-même une propriété plutôt fiable, stable dans le temps, et associée à d'autres mesures de façon cohérente. Elle paraît avoir même un composant génétique. Voir T. W. Britt, "METATRAITS - EVIDENCE RELEVANT TO THE VALIDITY OF THE CONSTRUCT AND ITS IMPLICATIONS," *Journal of Personality and Social Psychology* 65, no. 3 (1993), S. L. Hershberger, R. Plomin et N. L. Pedersen, "TRAITS AND METATRAITS - THEIR RELIABILITY, STABILITY, AND SHARED GENETIC INFLUENCE," *Journal of Personality and Social Psychology* 69, no. 4 (1995).

Cependant, ces facteurs ne sont apparemment pas associés à une augmentation du niveau de cohérence inter-situationnelle au-delà de limites qui sont incompatibles avec la notion de vertu traditionnelle.

Une bonne objection serait de remarquer que dans toutes ces expériences le groupe de départ est divisé en deux sous-groupes de la même dimension suivant la médiane : comme cela on a la moitié des gens qui possèdent le moins le méta-trait et la moitié des gens qui le possèdent le plus. Or, d'après la théorie de la vertu le sous-groupe des vertueux est vraisemblablement plus réduit que la moitié de la population. Dans ces conditions, les corrélations de la moitié des gens qui sont le plus vertueux seront inférieures à celle du groupe plus petit des gens vraiment vertueux. Toutefois, les corrélations du 50% des gens censés être les plus cohérents et à l'intérieur duquel on trouve le (disons) 20% des gens qui sont vraiment cohérents dans la population devraient être nettement plus élevées que les corrélations qu'on retrouve en général dans la population (sans division en sous-groupes), et cela n'est pas le cas, vu que le sous groupe des méta-traits semble lui-même être soumis au même plafond d'environ 0,30. Cela répond aussi à l'objection de Sabini : sans une référence directe au débat sur les méta-traits, mais envisageant la possibilité que seule une minorité des gens soient cohérents dans leur positionnement sur l'échelle d'un trait, il a soutenu qu'on aurait dû s'attendre aux coefficients de corrélation très peu élevés des études classiques, qui seraient donc compatibles avec l'existence de traits larges cohérents. Cela se produirait parce que l'incohérence du sous-groupe des incohérents fait tomber les corrélations globales du groupe par rapport aux bonnes corrélations qu'il y aurait dans le sous-groupe des cohérents³⁸². C'est exactement le même genre d'argument qui motive la stratégie des méta-traits. En réponse, Nous avons vu que des tentatives sophistiquées pour mesurer le niveau de cohérence du sous-groupe des personnes censées être les plus cohérentes ont pour l'instant montré que le plafond de Mischel résiste. Il ne sera jamais inutile de répéter que, d'un point de vue absolu, nos données empiriques ne sont pas idéales et ne permettent pas des conclusions au-delà de tout doute raisonnable. Mais il serait injustifié de les ignorer.

Conclusion

Depuis le début, dans les années 1920, les psychologues de la personnalité ont eu l'intention et l'ambition de conceptualiser et d'explorer empiriquement la possibilité de la cohérence comportementale impliquée par des traits larges, complexes et de haut niveau. Cela est prouvé par

³⁸² J. Sabini et M. Silver, "Lack of character? Situationism critiqued," *Ethics* 115, no. 3 (2005): 543.

le raffinement théorique des textes des pères fondateurs du domaine, tels que Allport ou Cattell. La pratique expérimentale, prise dans sa globalité (et si on ne se limite pas à analyser une expérience, l'œuvre d'un psychologue ou des tendances temporaires), a été suffisamment à la hauteur de cette ambition. Les psychologues ont essayé de dénicher la cohérence en explorant une série de possibilités bien plus large que celle prise en compte par les théoriciens de la vertu, ce qui est compréhensible, car, par exemple, ces derniers n'auraient pas accepté l'idée de traits entièrement subjectifs. Néanmoins, certaines hypothèses paraissent extrêmement pertinentes, par exemple l'idée que seule une minorité de la population serait cohérente et caractérisée par la perception de la centralité du trait concerné dans leur vie, de son caractère unique. L'échec dans la recherche de la cohérence sur une échelle aussi vaste va nous contraindre à élaborer une nouvelle conceptualisation des notions de caractère et de vertus dans la suite de ce travail.

Le menace radicale du situationnisme

Introduction au problème du pouvoir de la situation

Non seulement dans le cadre d'articles ou textes introductifs et de vulgarisation de psychologie sociale, mais de plus en plus en économie, jurisprudence et sciences humaines en général, on rencontre la thèse suivante : le comportement humain est en large partie (en tout cas, beaucoup plus qu'on ne le croirait) le fait d'influences de la situation. Qu'est-ce que cela veut dire exactement ? Commençons par une présentation intuitive : cela ne sera pas uniquement une stratégie d'exposition, car le message immédiat et saillant du situationnisme continue à jouer un rôle fondamental dans la recherche situationniste méthodologiquement raffinée (non simplement dans sa vulgarisation).

L'idée, qui normalement semble très intuitive au lecteur, est introduite par le biais d'une généralisation de cas paradigmatiques d'« influences situationnelles ». Si on vous braque un revolver sur la nuque, vous serez disposé à faire presque n'importe quoi, par exemple, donner au malfaiteur toutes les espèces que vous avez sur vous. Si vous êtes en train de mourir de faim, vous serez également disposé à faire peut-être pas n'importe quoi, mais toutes sortes de choses que vous ne feriez pas dans des conditions normales. Dans les deux cas, il s'agit de comportements influencés par la situation. La généralisation de ce genre de scénario, qu'elle soit plausible ou non, n'est certainement pas difficile à imaginer : il est possible que les gens fassent tout un tas de choses sous l'influence de facteurs situationnels moins évidents qu'un revolver braqué à la nuque ou l'étreinte de la faim.

Ces autres facteurs ne seraient pas seulement moins évidents mais souvent carrément cachés, insoupçonnés et anodins. C'est là tout l'intérêt d'un situationnisme généralisé : une influence situationnelle très saillante et pertinente invite déjà naturellement une explication situationniste. On n'aurait rien appris au sens commun s'il s'agissait de dire que la menace, la peur ou l'embarras éprouvés dans des situations extrêmes peuvent expliquer des choix qui vont à l'encontre des engagements et des désirs des acteurs. Un situationnisme qui ne soit pas au moins modérément ambitieux se réduit à un truisme. Il s'agit donc de dénicher l'influence des situations où on l'attend le moins.

Là non plus l'imagination n'a de limites : il est tout à fait facile de concevoir la possibilité que nous fassions beaucoup de choses sous la pression d'influences situationnelles cachées. Un engagement

important pris uniquement grâce à la bonne humeur ou à l'accent sympathique de la personne qui nous a demandé une faveur, un achat « forcé » par des aspects apparemment anodins mais efficaces d'une stratégie de marketing ou l'aide offerte avec sollicitude à quelqu'un qui a une malaise juste à côté de nous mais que nous aurions ignoré sans réfléchir si on s'était trouvé de l'autre côté de la route sont autant d'exemples. D'ailleurs, une démonstration expérimentale ne pose pas de problèmes spécifiques : il suffit de manipuler la présence de la situation étudiée en contrôlant tous les autres paramètres et de vérifier si le comportement est affecté d'une façon significative suivant que le sujet se retrouve ou non dans la toute-puissante situation concernée. Ainsi, si beaucoup plus de gens aident une personne qui a laissé tomber des papiers dans la rue après qu'ils ont trouvé de la monnaie dans un téléphone public, on dira que la bonne humeur inspirée par une variation anodine dans la situation influence l'altruisme³⁸³. Si le fait d'être légèrement pressé ou d'avoir le temps permet de prédire la disponibilité à aider une personne qui a un malaise, on dira également que l'altruisme a été affecté par des influences situationnelles³⁸⁴.

Le situationnisme défini intuitivement comme la thèse selon laquelle « le comportement humain est en large partie (en tout cas, beaucoup plus qu'on ne le croirait) le fait d'influences de la situation » paraît donc une position claire et méthodologiquement facile à évaluer. En tant que telle, en tout cas, elle est en train de conquérir l'attention d'économistes, théoriciens du droit et des sciences humaines plus généralement. Si la psychologie humaine est situationniste en ce sens, cela aurait des conséquences intéressantes et radicales dans tous ces domaines.

Or, le problème de cette nouvelle tendance est que le situationnisme ainsi défini est une position qui se révèle terriblement vague et équivoque. S'il était important de partir d'une présentation intuitive de la question (car l'attractivité intuitive du concept du « pouvoir de la situation » est le véritable fond de commerce de la mode situationniste), dès que l'on cherche à être plus précis, les problèmes ne se font pas attendre.

La thèse situationniste est le produit d'une tradition de recherche en psychologie sociale qui s'est développée depuis une quarantaine d'années³⁸⁵. La vaste sélection d'études empiriques qui figurent aujourd'hui dans la « bibliothèque situationniste » mise à jour sont dans leur ensemble le fondement

³⁸³ A. M. Isen et P. F. Levin, "EFFECT OF FEELING GOOD ON HELPING - COOKIES AND KINDNESS," *Journal of Personality and Social Psychology* 21, no. 3 (1972).

³⁸⁴ J. M. Darley et C. D. Batson, "FROM JERUSALEM TO JERICHO - STUDY OF SITUATIONAL AND DISPOSITIONAL VARIABLES IN HELPING BEHAVIOR," *Journal of Personality and Social Psychology* 27, no. 1 (1973).

³⁸⁵ Nous nous référons ici à ce que nous appelons tradition situationniste au sens strict, non pas à la conceptualisation et à la recherche de la cohérence du comportement que nous venons d'analyser dans le chapitre précédent.

d'une thèse de plus en plus conquérante dans le marché interdisciplinaire des idées : l'homme est un animal situationnel. Pour commencer, il est important de distinguer la littérature situationniste du situationnisme en tant que thèse. Une deuxième idée est qu'une littérature très vaste (qui pourrait avoir des conséquences intéressantes pour plusieurs raisons indépendantes) converge sur un message suffisamment clair et unitaire, en mesure de modifier profondément notre conception de la nature humaine. C'est bien ce message qui a des conséquences majeures, non pas les conclusions directes de chacune des études situationnistes : les études sont importantes parce qu'elles permettent dans leur ensemble d'étayer la thèse situationniste. Mais on pourrait très bien reconnaître l'importance de la littérature situationniste pour d'autres raisons et désavouer le situationnisme. Dans ce qui suit nous chercherons à présenter la littérature situationniste, à éclairer le situationnisme comme thèse et à déterminer dans quelle mesure il est soutenu par la littérature situationniste.

I. La littérature situationniste : la force de la situation et l'erreur fondamentale d'attribution

On a dit que la littérature situationniste au sens large est composée par la recherche de traits de caractère en psychologie de la personnalité et par la littérature situationniste proprement dite, c'est-à-dire la reproduction systématique d'effets qui prouvent le pouvoir de la situation. Cette dernière littérature est à son tour composée d'expériences censées prouver que les situations affectent le comportement et d'expériences qui cherchent à montrer que les gens ne se rendent pas compte de ce fait. Les deux genres d'expériences se recoupent partiellement, la différence étant la suivante : lorsqu'on vise l'attitude des gens qui ne se rendent pas compte du pouvoir de la situation, on s'intéresse à des cas où l'influence situationnelle paraît évidente aux expérimentateurs et aux lecteurs (p.ex. : on marque moins de paniers au basket si l'éclairage de la salle est mauvais) ; lorsqu'on vise le phénomène du pouvoir de la situation lui-même on s'intéresse plutôt à des cas où l'influence de la situation surprend tout le monde, expérimentateurs et lecteurs compris (ex : l'expérience de Milgram, dans laquelle environ deux tiers des participants acceptent d'électrocuter quelqu'un à 450 volts sous prétexte que le paradigme expérimental l'exige). La raison du recoupement n'est pas que des cas triviaux d'influence situationnelle où des gens avisés ne se méprennent pas peuvent être utiles pour étudier des aspects intéressants d'une psychologie situationnelle (car, nous l'avons dit, il y a rien à apprendre sur la pression situationnelle d'un revolver braqué sur la nuque) mais plutôt que les cas où les expérimentateurs sont également surpris

et aveugles au pouvoir de la situation sont intéressants pour étudier en général l'erreur de perception : il peut y avoir une continuité cognitive entre la cause d'erreurs flagrantes (ne pas comprendre qu'on met moins de paniers à cause du mauvais éclairage) et la cause d'erreurs plus subtiles (ne pas anticiper les résultats de Milgram). L'erreur des expérimentateurs serait alors un cas-limite des biais cognitifs des gens ordinaires.

Nous allons commencer par le premier genre de littérature situationniste proprement dite, l'étude du phénomène du pouvoir de la situation. Dans cette section nous allons mobiliser la notion d'« influence » et de « force » de la situation sans la clarifier ni la critiquer ; ce sera l'objet de la section suivante. Il sera nécessaire d'exposer dans les détails un bon nombre de paradigmes expérimentaux. Le lecteur pourra trouver cela pédant, la structure générale paraissant évidente après quelques exemples. Cependant, le caractère ouvert et illimité de la liste de situations exemplifiées par ces données est l'argument central des situationnistes : il est important d'en faire l'expérience.

En outre, les détails de chaque paradigme sont nécessaires pour l'interprétation situationniste. Dans une expérience typique de cette littérature, la preuve d'une influence insoupçonnée de la situation exige qu'on montre (a) qu'il n'y a pas de différences de trait de personnalité qui permettent de prédire le résultat mieux que la manipulation situationnelle, en d'autres termes, qu'il n'y a rien de plus général que la situation qui explique le résultat (b) que, une fois trouvée la bonne description des facteurs situationnels décisifs dans l'expérience, on puisse répliquer dans d'autres expériences presque identiques (ou très proches) le même effet du paradigme originel. Il ne faut pas faire la confusion entre l'idée qu'on pourrait découvrir partout des exemples du pouvoir de la situation – qui vont donc se révéler très hétérogènes dans leur manifestation et leur étiologie – et la précision nécessaire pour en établir l'existence dans un cas spécifique. Dès qu'on est convaincu, les exemples se multiplient, deviennent bon marché et les détails perdent de leur importance. Mais avant cela, pour avoir de bonnes raisons de se convaincre il faut passer par la réplification des détails et des effets de chaque paradigme.

Le paradigme de Milgram

Le paradigme expérimental de Milgram représente l'exemple le plus célèbre de la démonstration du pouvoir de la situation³⁸⁶. Au début des années 60 Stanley Milgram a conduit à Yale une longue série de variations et répétitions de l'expérience suivante : des sujets, dont la condition socioéconomique était très variée, ont répondu à des annonces ou ont été contactés par courrier pour

³⁸⁶ Stanley Milgram, *Obedience to authority; an experimental view*, 1ère édition. (New York: Harper & Row, 1974).

participer à une expérience sur la mémoire et l'apprentissage. Au laboratoire, chaque participant était accueilli par un expérimentateur en blouse blanche qui le présentait à un autre « participant » (en réalité un collaborateur) et l'informait que l'expérience porterait sur les effets de la punition sur l'apprentissage. Les rôles d'enseignant et d'élève étaient assignés par un faux tirage au sort. Après quoi le collaborateur (qui était toujours l'élève) était attaché à une chaise pour empêcher des mouvements excessifs, avec une électrode liée au poignet. Le participant était assuré que les décharges pouvaient être très douloureuses mais qu'elles ne pourraient pas endommager de façon permanente les tissus, il était aussi soumis à une faible décharge d'essai. Ensuite, il était conduit dans une autre pièce où il avait la possibilité d'entendre mais pas de voir l'élève et où se trouvait le générateur électrique. Le générateur avait 30 boutons : à partir de 15 volts on allait progressivement jusqu'à 450 volts, en montant de 15 volts à chaque fois. Les boutons étaient aussi accompagnés par une description : à 150 volts « strong shock », à 225 « intense shock », à 375 « Danger, sever shock » et à 435 et 450 tout simplement « xxx ». Le participant était invité par l'expérimentateur à actionner le dispositif en donnant à chaque fois une décharge de 15 volts plus intense pour toute erreur dans une tâche d'association de mots (et les erreurs ne se faisaient pas attendre, vu que l'élève était un collaborateur et les décharges n'étaient pas réelles). Le paradigme prévoyait que l'expérimentateur garde un ton ferme mais pas impoli, en réagissant de façon standardisée au refus éventuel du participants : d'abord « continuez s'il vous plaît » (*please continue*), après « l'expérience requiert que vous continuez » (*the experiment requires that you continue*), ensuite « il est absolument essentiel que vous continuez » (*it is absolutely essential that you continue*) et finalement « vous n'avez pas d'autre choix, vous devez continuer » (*you have no other choice, you must go on*). Si un épisode de refus persistait face à la quatrième et dernière réponse standardisée (l'expérimentateur avait aussi le droit de rappeler qu'il n'y aurait pas de dommage permanent aux tissus de l'élève), l'expérience était terminée et le sujet était considéré comme « non-obéissant ». Par contre, si – après un refus initial – le sujet décidait de continuer, la séquence reprenait depuis le début (et ainsi de suite, pour tout épisode de refus successif). Les participants qui allaient jusqu'au bout étaient classés comme « obéissants ». Les réactions de l'élève, que le participant pouvait entendre, était en fait enregistrées et suivaient aussi un schéma standardisé : à 150 volts (strong shock) on assistait déjà à une demande explicite d'interrompre l'expérience (se plaignant de problèmes cardiaques), à 270 il y avait des « cris agonisants » et un refus à continuer de plus en plus marqué, les cris se faisaient de plus en plus intenses et le refus répété de façon de plus en plus obsessive jusqu'à 330, de 345 à 450 il n'y avait aucune réaction de la part de l'élève (la séquence de décharge continuait parce que l'absence de réponse était considérée comme une erreur).

Dans le cadre des nombreuses répétitions et variations de l'expérience conduites par Milgram, aussi bien que dans les études qui ont successivement adopté son paradigme, environ deux tiers des participants sont obéissants et vont jusqu'au bout, tout en montrant de nombreux signes de détresse, angoisse et anxiété : ils transpiraient, tremblaient, bégayaient, mordaient leurs lèvres, grognaient, enfonçaient leur ongles dans la chair³⁸⁷. Ce résultat est bien sûr étonnant pour une série de raisons, mais à ce stade nous allons uniquement nous intéresser à ce qui fait de ce paradigme un classique situationniste (et non pas, par exemple, une démonstration de l'existence d'une propension générale à la soumission dans la population)³⁸⁸.

La démonstration du pouvoir de la situation réside dans la grande variation dans le taux d'obéissance qui a été produite en modifiant des aspects apparemment mineurs du paradigme, par opposition à capacité modeste de différences au niveau de traits de prédire le comportement des participants. La variation du sexe et de l'âge moyen du groupe n'est pas associée à un taux

³⁸⁷ S. Milgram, "BEHAVIORAL-STUDY OF OBEDIENCE," *Journal of Abnormal Psychology* 67, no. 4 (1963). Pour des raisons éthiques évidentes, le paradigme de Milgram est plutôt problématique, ce qui a limité le nombre des répétitions, dont la plupart a eu lieu avant la moitié des années 80, lorsque les normes éthiques pour la recherche étaient plus relaxées.

³⁸⁸ En ce qui concerne la validité écologique, les critiques se sont concentrées sur deux objections majeures. D'un côté, on peut considérer que l'expérience n'est pas significative car, après tout, les sujets pouvaient se sentir pleinement justifiés en faisant une confiance totale à l'expérimentateur. L'autorité d'un scientifique de Yale paraît suffisante pour autoriser un transfert complet de responsabilité. Si l'expérimentateur dit que tout va bien cela doit être vrai. Cette objection ne paraît pas très convaincante d'un point de vue normatif, car le participant entendait clairement l'élève crier et demander que l'expérience soit interrompue et on ne voit pas quel genre d'autorité pourrait assurer que tout va bien face à un refus explicite. D'ailleurs, il y a des répliques où l'expérience ne se fait pas à Yale mais dans un entrepôt en banlieue, ou dans lesquelles l'expérimentateur n'assure pas qu'il n'y aura pas de dommage aux tissus, ou encore il se montre très surpris et incertain face aux protestations de l'élève (mais il décide que l'expérience doit continuer) : le taux d'obéissance ne change pas beaucoup. La deuxième objection pose plus de problèmes : il s'agit du fait que les participants pouvaient douter de la réalité des décharges électriques. Dans un questionnaire envoyé par Milgram un an après l'expérience 80% des participants qui ont répondu affirmaient être convaincus de la réalité des décharges (Milgram 74 : 172-3) et Doris (44) soutient qu'un doute léger ne justifie pas l'obéissance d'un point de vue éthique. Cependant, la question de savoir si le taux d'obéissance serait le même dans un contexte réel où il n'y a pas de doute possible reste ouverte : même si le doute ne constitue pas une justification il pourrait bien affecter et expliquer le comportement des gens. On ne sait pas si Milgram a découvert le même phénomène psychologique qui explique la collaboration des gens normaux à des massacres ou à des génocides.

d'obéissance significativement différent, ainsi que la nationalité³⁸⁹. Egalement, des mesures de traits de personnalité comme l'échelle F pour la personnalité autoritaire développée par Adorno, le niveau de maturité du raisonnement moral dans l'échelle de Kohlberg et d'autres mesures répandues à l'époque n'ont que des corrélations modestes avec l'obéissance dans le paradigme de Milgram³⁹⁰. En revanche, il suffit de faire en sorte que l'expérimentateur soit physiquement absent et donne ses ordres par téléphone pour faire baisser le taux d'obéissance à 21% ; si le participant a le choix du voltage des décharges, seuls 3% arrivent à administrer la décharge maximale ; si l'enseignant doit pousser physiquement la main de l'élève sur l'électrode l'obéissance se réduit à 30% et elle est presque nulle s'il y a deux expérimentateurs en désaccord quant à l'opportunité de continuer, par contre si le participant ne doit pas administrer lui-même les décharges mais qu'il se limite à participer à l'expérience avec des tâches secondaires, le taux remonte à 91%. Des résultats de ce genre constituent des cas paradigmatiques d'influence de la situation : des variations dans le contexte produisent des effets inattendus au niveau du comportement et qui ne paraissent ni justifiés ni proportionnés au changement qui les a déterminés.

Les effets sur l'humeur

Une série d'expériences situationnistes ont étudié les effets sur le comportement de légers changements d'humeur dans l'attitude des gens. Dans une expérience célèbre³⁹¹, un collaborateur faisait tomber exprès des papiers dans la rue près de sujets qui venaient de sortir d'une cabine téléphonique où ils avaient ou non trouvé de la monnaie. Le but était de voir si les participants aidaient ou non dans les deux cas. Le résultat a été le suivant :

	Aide (nombre de sujets)	N'aide pas
Monnaie	14	2
Pas de monnaie	1	24

³⁸⁹ Voir D. M. Mantell, "POTENTIAL FOR VIOLENCE IN GERMANY," *Journal of Social Issues* 27, no. 4 (1971).

Occasionnellement, on peut trouver des taux d'obéissance plus importants – 77% pour les Allemands (1971), mais deux tiers d'obéissants est le résultat typique.

³⁹⁰ T. Blass, "UNDERSTANDING BEHAVIOR IN THE MILGRAM OBEDIENCE EXPERIMENT - THE ROLE OF PERSONALITY, SITUATIONS, AND THEIR INTERACTIONS," *Journal of Personality and Social Psychology* 60, no. 3 (1991).

³⁹¹ Isen et Levin, "EFFECT OF FEELING GOOD ON HELPING - COOKIES AND KINDNESS."

Dans ce cas, les participants n'ont pas été soumis à des tests de personnalité et la supériorité de l'influence situationnelle sur des prédispositions individuelles à l'altruisme ne peut être documentée. Cependant, vu la différence aussi nette entre les deux groupes, il serait ridicule de supposer que, par un hasard statistique, le deuxième groupe serait composé par des individus méchants : l'influence du facteur situationnel est évidente.

Baron et Thomley ont montré que la probabilité que des gens aident dans des tâches telles que ramasser un stylo que quelqu'un (du même sexe) a laissé tomber ou faire de la monnaie d'un billet d'un dollar double (d'environ 20-25% à 50-60% de gens qui aident) en présence d'une odeur agréable de boulangerie ou de café³⁹². Les gens sont disposés à offrir plus que le double de leur temps (de dix minute à entre 20 et 30) pour aider l'expérimentateur sans récompense après la fin de l'expérience s'ils ont reçu un petit cadeau ou qu'il y a un parfum de citron ou de fleurs dans la pièce³⁹³. En présence d'un bruit très fort produit par une tondeuse le pourcentage de gens qui aident un garçon au bras bandé qui fait tomber une énorme pile de livres passe de 80% à 15%³⁹⁴. Effets semblables quoique souvent d'une ampleur inférieure ont été attribués à de la musique d'ambiance agréable³⁹⁵, au soleil, à l'humidité, température et vitesse du vent³⁹⁶.

Dans ces expériences on s'étonne moins que le pouvoir de prédiction des facteurs individuels ne soit pas décisif³⁹⁷ : cela se justifie par le fait que l'existence de ces facteurs situationnels qui

³⁹² R. A. Baron, "The sweet smell of ... helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls," *Personality and Social Psychology Bulletin* 23, no. 5 (1997).

³⁹³ R. A. Baron et J. Thomley, "A WHIFF OF REALITY - POSITIVE AFFECT AS A POTENTIAL MEDIATOR OF THE EFFECTS OF PLEASANT FRAGRANCES ON TASK-PERFORMANCE AND HELPING," *Environment and Behavior* 26, no. 6 (1994).

³⁹⁴ K. E. Mathews et L. K. Canon, "ENVIRONMENTAL NOISE-LEVEL AS A DETERMINANT OF HELPING BEHAVIOR," *Journal of Personality and Social Psychology* 32, no. 4 (1975).

³⁹⁵ R. Fried et L. Berkowitz, "MUSIC HATH CHARMS ... AND CAN INFLUENCE HELPFULNESS," *Journal of Applied Social Psychology* 9, no. 3 (1979).

³⁹⁶ M. R. Cunningham, "WEATHER, MOOD, AND HELPING-BEHAVIOR - QUASI EXPERIMENTS WITH THE SUNSHINE SAMARITAN," *Journal of Personality and Social Psychology* 37, no. 11 (1979). Inversement, une chaleur excessive semble être liée à des épisodes de violence, voir C. A. Anderson et al., "Temperature and aggression," in *Advances in Experimental Social Psychology, Vol 32, Advances in Experimental Social Psychology* (San Diego: Academic Press Inc, 2000).

³⁹⁷ Dans cette littérature il y a des résultats contradictoires sur le pouvoir de prédiction de variables démographiques telles que le genre ou le statut socio-économique (ex. : parfois les hommes sont plus altruistes, parfois moins), négatifs pour des traits classiques en psychologie de la personnalité (ex : personnalité autoritaire, autonomie, Machiavélisme) et faibles pour des facteurs tels que la « tolérance pour l'ambiguïté » et la « tendance à accepter la responsabilité », voir Doris, *Lack of character: personality and moral behavior* 36-38.

influencent l'humeur est déjà reconnue par le sens commun, il n'y a pas besoin de convaincre qu'il s'agit d'influence de la situation plutôt que de variabilité interpersonnelle. Tout l'intérêt réside dans l'ampleur de l'effet, supérieure à ce que les gens s'attendent : personne ne s'étonne du fait qu'être dans un bon état d'esprit pour des bonnes raisons puisse affecter notre disposition à aider autrui ; le point est plutôt qu'une variation minimale d'humeur induite par des facteurs anodins est en mesure d'exercer une influence disproportionnée sur nos choix éthiques.

Le paradigme du Bon Samaritain

Une expérience très célèbre est celle conduite par Darley et Batson (et après répliquée maintes fois) dans le cadre d'un séminaire de théologie à Princeton en 1973. Les séminaristes ont été invités à participer à une étude sur le thème « éducation religieuse et vocations » en échange de 2,5 dollars. Dans une première phase, on leur demandait de remplir des formulaires dans un bâtiment pour après se déplacer dans un autre où ils devaient faire une courte présentation orale. Au moment de quitter le premier bâtiment, ils étaient prévenus qu'ils étaient soit très en retard, soit qu'ils avaient juste le temps d'y aller ou encore qu'ils étaient un peu en avance. On s'intéressait à voir si les sujets se seraient arrêtés pour porter secours à un sujet en détresse (en réalité un collaborateur) semi-allongé dans une porte par laquelle il fallait passer pour se rendre au deuxième site. Les résultats ont été les suivants :

	Pas pressé	Moyennement pressé	Très pressé
% aide	63	45	10

Il apparaît que le fait d'être pressé dans ce genre de situation est en mesure d'exercer une influence considérable sur le comportement. Pour la moitié des sujets la présentation qu'ils devaient faire portait sur la parabole du Bon Samaritain (d'où le nom du paradigme expérimental, qui n'est pas uniquement inspiré par la disposition à aider directement testée), les autres avaient un sujet moins pertinent : la différence n'a pas affecté de façon significative le résultat. Des mesures de traits de personnalité censées correspondre à des styles de religiosité différents n'ont pas été utiles pour prédire le comportement des séminaristes³⁹⁸.

³⁹⁸ Darley et Batson, "FROM JERUSALEM TO JERICHO - STUDY OF SITUATIONAL AND DISPOSITIONAL VARIABLES IN HELPING BEHAVIOR," 106.

L'influence du groupe

Une grande quantité d'expériences ont montré la différence d'attitude des gens selon qu'ils sont seuls ou dans un groupe : avec un ou plusieurs partenaires – connus ou inconnus – on ne se comporte pas de la même façon, la différence n'allant pas d'ailleurs dans le même sens dans ces quatre cas de figure. Encore une fois, rien d'exceptionnel ou de nouveau, si ce n'est l'influence disproportionnée de variations apparemment anodines dans la composition et dans l'attitude de notre entourage social. Latané et Darley ont fait rentrer de la fumée lentement par une buse d'aération dans une pièce où des étudiants étaient en train de remplir des formulaires³⁹⁹. Au bout d'un certain temps, le niveau de la fumée était suffisant pour empêcher de voir et respirer correctement, il y avait aussi une odeur désagréable. Le but était de voir dans quelles conditions les étudiants auraient prévenu l'expérimentateur du problème, le résultat a été le suivant :

	Seul	Avec deux collaborateurs (anonymes) qui font mine de rien	Groupe de 3 participants
% étudiants qui préviennent	75%	10%	38%

Dans une autre expérience, une jeune femme séduisante donnait des formulaires à remplir à des étudiants pour une recherche de marché et après se retirait derrière une cloison dans la pièce, hors de vue. Ici un enregistrement simulait de façon très réaliste (presque personne n'a signalé après d'avoir eu des soupçons) le bruit et les cris de douleur d'une chute⁴⁰⁰. Les pourcentages des étudiants qui sont intervenus pour aider sont les suivants :

	Seul	Avec collaborateur qui fait mine de rien	2 participants
% intervention	70%	7%	40%

Dans une expérience où les participants sont seuls dans une cabine mais croient être en contact par interphone (pour préserver l'anonymat) soit avec une seule personne, soit avec un groupe de

³⁹⁹ Bibb Latané et John M. Darley, *The unresponsive bystander: why doesn't he help?*, Century psychology series (New York, : Appleton-Century Crofts, 1970) 44-54.

⁴⁰⁰ B. Latane et J. Rodin, "A LADY IN DISTRESS - INHIBITING EFFECTS OF FRIENDS AND STRANGERS ON BYSTANDER INTERVENTION," *Journal of Experimental Social Psychology* 5, no. 2 (1969).

discussion de 5 personnes, lorsque l'individu à l'autre bout de l'interphone simule une attaque épileptique (il s'agissait en réalité d'un enregistrement) 100% des sujet qui se croient seuls interviennent, alors que seulement 62% des participants « en groupe » le font⁴⁰¹.

Le paradigme de Asch sur les influences du groupe sur la cognition est très célèbre⁴⁰². Asch a demandé à des participants de juger laquelle de trois lignes affichées était de la même longueur qu'une quatrième ligne de référence. La tâche était extrêmement facile et le taux d'erreur d'un sujet seul presque négligeable. En groupe avec six collaborateurs qui donnaient exprès des réponses incorrectes, au fil de plusieurs essais environ 75% des sujets « cèdent au groupe » au moins une fois (tout de même, 75% donnent la réponse incorrecte moins qu'une fois sur quatre). L'interprétation de cette expérience a été récemment contestée : traditionnellement, on suppose que d'un point de vue normatif les sujets devraient donner toujours la réponse correcte, toute erreur paraît l'effet inattendu et injustifié d'une influence de la situation et doit être expliquée par des mécanismes d'altération de la perception ou par l'embarras face au groupe. Hodges et Geyer pensent que ces attentes normatives ne sont pas défendables : dans une approche pragmatique qui se concentre sur l'interaction à long terme il paraît raisonnable de pas dire la vérité parfois pour gérer les tensions dans le groupe⁴⁰³. S'ils ont raison, les expériences de Asch ne constitueraient des preuves du pouvoir de la situation plus intéressantes que celles normalement reconnues par le sens commun (ex : le pouvoir situationnel d'un revolver braqué sur la nuque), car on *devrait* s'attendre à ce que les gens réagissent comme ils le font dans le laboratoire.

Influences positives de la situation : altruisme

Malgré le biais général en psychologie pour les aspects négatifs et les problèmes du comportement humain (et la psychologie sociale n'est certainement pas une exception !), il ne manque pas de preuve d'effets positifs des situations⁴⁰⁴. Dans une expérience conduite en 1972⁴⁰⁵ avec des

⁴⁰¹ Latané et Darley, *The unresponsive bystander: why doesn't he help?* 95-100.

⁴⁰² S. E. Asch, "STUDIES OF INDEPENDENCE AND CONFORMITY .1. A MINORITY OF ONE AGAINST A UNANIMOUS MAJORITY," *Psychological Monographs* 70, no. 9 (1956).

⁴⁰³ B. H. Hodges et A. L. Geyer, "A nonconformist account of the Asch experiments: Values, pragmatics, and moral dilemmas," *Personality and Social Psychology Review* 10, no. 1 (2006).

⁴⁰⁴ Pour une analyse très détaillée de cette littérature, voir Peter B. M. Vranas, "The indeterminacy paradox: character evaluations and human psychology" (non publié: The University of Michigan, 2004).

⁴⁰⁵ R. D. Clark et L. E. Word, "WHERE IS APATHETIC BYSTANDER - SITUATIONAL CHARACTERISTICS OF EMERGENCY," *Journal of Personality and Social Psychology* 29, no. 3 (1974).

étudiants (de sexe masculin) contactés au hasard et payés deux dollars pour participer à une étude de validation d'un test pour évaluer les étudiants de première année, l'expérimentateur accompagnait chaque sujet vers le site de l'expérience, en passant par une pièce où un technicien, au milieu de câbles et de signes de danger pour haute tension très visibles, était apparemment en train de faire de réparations. Soudainement le sujet voyait un éclair de lumière, entendait un bruit sourd de courant et le cri du technicien qui tombait en restant bien emmêlé dans les câbles. Presque tous les participants ont aidé le technicien (une seule exception), mais, ce qui est étonnant, 71% l'ont aidé directement, en le détachant des câbles avec leurs mains, malgré le danger évident. 65% ont admis ne pas avoir les compétences pour être sûrs de ne pas être électrocutés dans l'intervention. Le résultat a été répliqué mais uniquement avec des étudiants de sexe masculin⁴⁰⁶.

Plusieurs expériences montrent que les gens sont disposés à poursuivre et arrêter un voleur si un inconnu leur demande de faire attention à leurs affaires. Dans les expériences de Moriarty une personne range sa serviette à côté du participant (involontaire, dans ce cas) à la plage, s'allonge, écoute de la musique pendant quelques minutes et après lui demande qu'il fasse attention à ses affaires (dont la radio) pendant quelque temps. La scène peut aussi se dérouler dans une cafétéria de Manhattan où les victimes sont des personnes qui dînent seules : un collaborateur s'assoit à la même table et demande qu'on fasse attention à son sac pendant quelques minutes. Dans les deux scénarios un (faux) voleur arrive, prend les affaires ou le sac et s'enfuit⁴⁰⁷. Il se trouve que tous les participants (10 pour la plage et 8 dans la cafétéria) dans ces deux expériences ont poursuivi et cherché à attraper le voleur ! Des résultats aussi impressionnants ont été répliqués avec un échantillon très nombreux (il s'agissait de faire attention à une calculatrice et à un cahier dans un couloir d'université) : bien 76% des participants sont intervenus⁴⁰⁸. Celle-ci et d'autres répliques ont prouvé que le résultat est significativement stable pour toutes les permutations de sexe (participant garçon, propriétaire des affaires fille, voleur garçon ; fille/garçon/fille ; garçon/garçon/fille etc.) et l'âge des participants a été varié aussi⁴⁰⁹. Le pourcentage ne passe pas au dessous de 65-70% même si le voleur est moyennement ou très costaud⁴¹⁰.

⁴⁰⁶ Ibid.: 280-283.

⁴⁰⁷ T. Moriarty, "CRIME, COMMITMENT, AND RESPONSIVE BYSTANDER - 2 FIELD EXPERIMENTS," *Journal of Personality and Social Psychology* 31, no. 2 (1975).

⁴⁰⁸ W. Austin, "SEX-DIFFERENCES IN BYSTANDER INTERVENTION IN A THEFT," *Journal of Personality and Social Psychology* 37, no. 11 (1979).

⁴⁰⁹ L. Schwarz et al., "ROLE OF COMMITMENTS IN THE DECISION TO STOP A THEFT," *Journal of Social Psychology* 110, no. 2 (1980). D. R. Shaffer, M. Rogel et C. Hendrick, "INTERVENTION IN LIBRARY -

Un autre paradigme est celui du viol simulé, où la plupart des sujets essaient d'arrêter un viol. 65% des sujets seuls interviennent directement plutôt qu'ignorer le fait ou prévenir la police (la simulation avait lieu relativement près du bâtiment de la police dans un campus universitaire) par rapport à 85% des participants qui marchaient dans un petit groupe de 3 à 5 personnes⁴¹¹. Il est intéressant de remarquer que l'effet du groupe peut être aussi positif, contrairement aux expériences de dilution de la responsabilité. Encore une fois, cela soutient l'approche situationniste qui attribue le pouvoir explicatif à la situation spécifique (décrite au niveau de généralité correct), non pas à des tendances plus générales. Que l'influence du groupe n'aille pas toujours dans le même sens mais dans des directions qu'on peut reproduire fidèlement dans chaque paradigme particulier marque un point pour le situationniste.

Dissonance cognitive et « channeling factors »

Dans une expérience célèbre, Festinger et Carlsmith⁴¹² ont occupé pendant une longue période des étudiants avec des tâches très faciles et extrêmement ennuyeuse que visiblement ils n'aimaient pas. A la fin, l'expérience prévoyait trois conditions : à un sous-ensemble de sujets on demandait en échange de 20 dollars la faveur de convaincre quand même une autre personne à participer, à un deuxième sous-ensemble on demandait la même chose en offrant un dollar et un troisième de contrôle était laissé libre de partir. Au bout d'un certain temps, on leur demandait d'évaluer l'intérêt de la tâche et il se trouve que le groupe qui avait été payé un dollar la trouvait sensiblement moins ennuyante que les deux autres.

Ce résultat a été expliqué par le mécanisme célèbre de la dissonance cognitive, l'idée étant que, comme il est trop difficile de justifier à ses propres yeux d'avoir menti pour un dollar, le sujet finit par se convaincre qu'après tout la tâche n'était pas si ennuyeuse. Ce genre d'explication a marqué l'histoire de la psychologie et du marketing et a été défendue ou attaquée par plusieurs arguments dans le cadre d'un grand nombre de paradigmes expérimentaux différents. Quoi qu'il en soit du

EFFECT OF INCREASED RESPONSIBILITY ON BYSTANDERS WILLINGNESS TO PREVENT A THEFT," *Journal of Applied Social Psychology* 5, no. 4 (1975).

⁴¹⁰ Austin, "SEX-DIFFERENCES IN BYSTANDER INTERVENTION IN A THEFT," 2115.

⁴¹¹ H. Harari, O. Harari et R. V. White, "THE REACTION TO RAPE BY AMERICAN MALE BYSTANDERS," *Journal of Social Psychology* 125, no. 5 (1985). Voir aussi Jane Allyn Piliavin, *Emergency intervention* (New York: Academic Press, 1981).

⁴¹² L. Festinger et J. M. Carlsmith, "COGNITIVE CONSEQUENCES OF FORCED COMPLIANCE," *Journal of Abnormal and Social Psychology* 58, no. 2 (1959).

mécanisme général de la dissonance cognitive, et même si cette explication n'était acceptable, l'expérience décrite illustre parfaitement l'influence de la situation : être payé un dollar plutôt que 20 pour faire la même chose dans le même contexte produit une différence inattendue et injustifiée. La différence entre 1 et 20 dollars n'était certainement pas anodine à l'époque mais le paradigme a été répliqué après avec des sommes inférieures.

Il est intéressant de remarquer que l'effet peut être considéré comme neutre d'un point de vue moral : formellement il s'agit de mentir, mais un mensonge dit pour faire en sorte qu'une expérience puisse se dérouler n'est pas problématique d'un point de vue éthique. En fait, la liste d'exemples qu'on vient d'analyser est très biaisée par rapport à la pertinence morale. Les exemples qui attirent le plus l'attention sont liés à des influences situationnelles qui nous font agir de façon étonnamment bonne ou mauvaise, mais le pouvoir de la situation est présent d'une manière beaucoup plus générale. Les exemples d'influence neutre où les gens sont amenés par la situation à faire quelque chose qu'ils n'auraient pas fait (sans qu'aucune des deux options ne soit spécialement condamnable ou louable : par exemple, acheter un produit plutôt qu'un autre, ou acheter plutôt que ne pas acheter) se multiplient et sont au centre de décennies de recherche en marketing, entre autres. Kurt Lewin, l'un des fondateurs de la psychologie sociale contemporaine (après avoir joué un rôle dans la *Gestalttheorie*), avait introduit dès les années 1930 un concept en mesure de couvrir la plupart de ces cas, celui de « *channeling factor* »⁴¹³. Un *channeling factor* est un facteur situationnel qui permet à une attitude de se concrétiser en action, il s'agit de ce à quoi on se réfère lorsqu'on dit « j'aurais toujours voulu le faire, mais je n'ai jamais trouvé l'occasion ou le contexte ».

Il y a une infinité de manipulations concevables en mesure de bloquer ou canaliser les attitudes des gens, comme tout architecte ou urbaniste le sait très bien. En changeant les dimensions ou la disposition des places assises dans une pièce, il y aura plus ou moins de personnes qui trouveront le courage d'intervenir et poser des questions. Ces effets moralement neutres sont beaucoup plus difficilement prévisibles que l'on ne croit. Certes, il y en a de triviaux que le sens commun connaît très bien (ex : je n'ai pas posé de questions car il y avait trop de monde), mais là encore il ne faut pas se tromper : nous sommes bien avisés et notre répertoire explicatif est très riche, on sait très bien que si je n'arrive pas à aborder une fille dans un bar cela peut être juste parce que je n'ai pas trouvé le bon contexte ou un prétexte adéquat, mais on s'étonnerait de découvrir l'effet *spécifique* d'un changement donné dans la situation, par exemple, la disposition des tables. Une quantité indéfinie d'exemples de *channeling factors* correspondent à la définition d'une preuve typique du

⁴¹³ Kurt Lewin, Donald Keith Adams et Karl Edward Zener, *A dynamic theory of personality; selected papers*, 1st édition., McGraw-Hill publications in psychology (New York, London, : McGraw-Hill Book Company inc., 1935).

pouvoir de la situation : l'effet est surprenant, répliquable (dans une description suffisamment détaillée de la situation) et aucune variable personnelle plus générale ne prédit mieux le résultat. Il ne faudra pas oublier cela même si nous nous intéressons avant tout au cas de la morale, car, comme nous le verrons, le vrai argument des situationnistes contre la théorie de la vertu dépend d'une généralisation très forte du pouvoir de la situation.

L'erreur fondamentale d'attribution

Une partie importante de la littérature situationniste est consacrée à montrer non pas le pouvoir de la situation, mais le fait que les gens l'ignorent ou le sous-estiment. Ces études font partie d'une littérature plus vaste et générale sur l'attribution de traits, attitudes et croyances par le sens commun : en psychologie sociale, on parle de « théorie de l'attribution ». Comme nous l'avons fait jusqu'à maintenant, ici nous nous limiterons à montrer pourquoi ces expériences ont été interprétées *prima facie* comme signes de l'ignorance répandue sur le pouvoir de la situation ; nous chercherons à dépasser une compréhension intuitive de ce phénomène après.

Pour commencer, toutes les expériences situationnistes dont nous venons de parler sont a fortiori des preuves du fait que les gens sous-estiment la force de la situation, car non seulement la personne commune, mais l'expérimentateur n'a su non plus prédire de tels résultats.

Toutefois, une longue tradition expérimentale a étudié les effets plus extrêmes du même phénomène, où les gens se trompent dans des cas qui paraissent normativement évidents aux expérimentateurs. Cette tradition de recherche a commencé avec l'étude célèbre de John et Harris sur l'attribution d'attitudes⁴¹⁴. On demandait à des étudiants d'évaluer l'attitude pro-Castro d'autres étudiants à partir d'essais que ces derniers avaient écrits. Les évaluateurs savaient que le choix de la ligne politique de l'essai n'était pas libre : les expérimentateurs avaient demandé d'écrire soit un texte pro-Castro, soit un texte anti-Castro, indépendamment des opinions individuelles. Il se trouve quand même que les auteurs de textes pro-Castro ont été évalués comme étant sensiblement plus pro-Castro que les autres (mais moins, bien sûr, que quelqu'un qui aurait écrit le même essai sans constrictions). Les évaluateurs considéraient les auteurs des textes pro-Castro comme étant *vraiment* pro-Castro, même si le choix de la thèse à défendre dans l'essai était imposé.

L'effet est considéré comme un cas paradigmatique de la tendance à ignorer la force de la situation pour se centrer sur les caractéristiques distinctives d'une personne. Comme il s'agit d'une attribution hâtive, il a été nommé « erreur fondamentale d'attribution » (*fundamental attribution*

⁴¹⁴ Edward E. Jones et Victor A. Harris, "The attribution of attitudes," *Journal of Experimental Social Psychology* 3, no. 1 (1967).

error)⁴¹⁵. S'il paraît étonnant que les gens attribuent des attitudes avec cette aisance, les surprises ne sont pas terminées. Chez des sujets américains, l'effet est réduit mais pas éliminé dans une variation où les évaluateurs font préalablement l'expérience d'écrire eux-mêmes un essai sous contrainte. L'effet ne disparaît pas non plus si on explique aux évaluateurs que les étudiants qui ont écrit les textes ont juste fait un copier-coller à partir d'arguments déjà développés (toujours en n'ayant pas choisi la ligne politique)⁴¹⁶ ! Dans la même expérience, cette persistance du biais est absente chez des étudiants coréens.

Dans la littérature interculturelle, ce qui émerge de dizaines d'études est que, si dans certaines situations spécifiques (celles où l'attention des gens n'est pas focalisée ou même explicitement écartée du rôle de la situation) tout le monde montre un niveau non nul du biais (mais pas toujours le même), plusieurs variables modulent la capacité de tenir en compte de la situation dans des paradigmes où son influence est soulignée par des manipulations expérimentales. Des gens dans des cultures moins individualistes, des personnes qui ne manifestent pas l'attitude individualiste de leur culture⁴¹⁷ et ceux qui croient que la personnalité peut changer et évoluer au fil du temps⁴¹⁸ sont en mesure de profiter de ces manipulations pour corriger le biais. Les occidentaux qui partagent l'attitude individualiste de leur culture, les individualistes dans les cultures collectivistes et ceux qui croient que la personnalité est fixée une fois pour toutes maintiennent un niveau non nul du biais même dans ces conditions.

II. Genres de « situationnismes » et « situations » différents en philosophie et psychologie

⁴¹⁵ Il est connu aussi sur d'autres dénominations, par exemple biais de sur-attribution (*overattribution bias*), biais de correspondance (*correspondence bias*) ou dispositionnisme populaire (*lay dispositionsim*), voir John Sabini, Michael Siepmann et Julia Stein, "The really fundamental attribution error in social psychological research," *Psychological Inquiry* 12, no. 1 (2001).

⁴¹⁶ Incheol Choi et Richard E. Nisbett, "Situational salience and cultural differences in the correspondence bias and actor-observer bias," *Personality and Social Psychology Bulletin* 24, no. 9 (1998).

⁴¹⁷ On parle de « allocentriques » pour se référer aux individus qui font exception dans une majorité de personnes « idiocentriques » (individualistes). La distinction « allocentriques/idiocentriques » est l'équivalent au niveau individuel de la distinction « collectiviste/individualiste » référée à des cultures.

⁴¹⁸ Il s'agit de la distinction entre « incremental » (malléable, ouverte) et « entity theory of personality » (fixe) étudiée par Carol Dweck, voir Daniel C. Molden et Carol S. Dweck, "Finding "meaning" in psychology - A lay theories approach to self-regulation, social perception, and social development," *American Psychologist* 61, no. 3 (2006).

Le situationnisme par définition semble donner un rôle essentiel à la notion de situation, en particulier à la « force de la situation » dans le comportement humain. Pour comprendre ce que cela pourrait vouloir dire, il sera utile de comparer le situationnisme en psychologie sociale à d'autres théories qui ont donné un rôle central à la situation, en psychologie et en philosophie.

Au moins dans son acception ordinaire et intuitive, la situation dans laquelle se trouve un individu joue clairement un rôle majeur dans l'explication de son comportement. Par « situation » à laquelle est confronté un individu au moment t on entend un sous-ensemble approprié de l'état global du monde au moment t formé par tout ce qui est en mesure d'affecter dans un sens quelconque l'individu. La situation peut être définie d'une manière objective, en excluant les états mentaux de l'individu (par exemple dans un usage comme « quand on est à la poste, il faut.. »), subjective, en excluant l'état spécifique du monde extérieur (ex. : « lorsqu'on est déprimé, il est important de... ») ou objective et subjective (ex. : « lorsqu'on est déprimé dans une queue à la poste, il faut... »). Ces notions de situation sont extrêmement vagues et leurs contours sont mal définis : on peut être en même temps en un nombre indéterminé de situations objectives, subjectives et objective-subjective. Par exemple, je peux être en même temps dans les situations suivantes : « à la poste », « à la poste dans une queue », « à la poste du 4^{ème} arrondissement de Paris », « en queue à la poste en train de déprimer », « en train de déprimer » etc. Tant qu'il n'y a pas d'ambition explicative exagérée, le vague de la notion courante de situation n'est pas un problème, et il est clair que les situations en ce sens jouent un rôle essentiel dans la production du comportement. L'action est le résultat de l'interaction entre l'environnement et l'individu, et l'état d'un sous-ensemble du système individu-environnement est bien sûr pertinent pour l'expliquer. Ce n'est donc pas dans ce sens trivial que la situation peut jouer un rôle majeur (ou être toute-puissante).

Une notion informative et substantielle de situation peut être déployée à trois niveaux : ontologique, psychologique et normatif. Au niveau ontologique, on peut défendre la nécessité d'un découpage objectif de l'état du monde en situations, au sens où certaines descriptions de la situation dans laquelle le sujet se trouve auraient des propriétés spéciales qui doivent être reconnues en tant que telles par le sujet. Par exemple, les particularistes moraux soutiennent que certains aspects du contexte nous donnent des raisons objectives d'agir qui ne peuvent pas être dérivées par l'application de principes généraux et qui doivent être à chaque fois reconnues par l'agent moral⁴¹⁹. Si je suis en train de me promener avec un ami et que tout à coup son attitude rend appropriée une proposition d'aide et de soutien de ma part, j'ai beau me trouver en même temps dans quantité de situations (au sens intuitif) qui pourraient attirer mon attention et occuper mes ressources cognitives

⁴¹⁹ Jonathan Dancy, *Ethics without principles* (Oxford ; New York: Clarendon, 2004).

(ex. : je suis fatigué, il fait beau, j'ai un autre rendez-vous, je suis en retard) : il faut que je reconnaisse l'objectivité de la situation morale. Dans ce contexte spécifique, l'attitude de mon ami est un fait moral objectif que je suis tenu de reconnaître. Le particularisme défend une ontologie où il y a une place non éliminable pour une notion informative de situation, et il défend aussi un rôle non éliminable des situations dans une théorie normative, car on ne peut pas déterminer ce que l'agent devrait faire par des principes généraux, sans faire référence au statut normatif de situations morales spécifiques.

Dans un autre sens encore, une notion intéressante de situation peut jouer un rôle crucial dans une théorie psychologique. C'est le cas par exemple du béhaviorisme et de la théorie du contrôle par le stimulus.

Si le système cognitif est sensible à de nombreux aspects du contexte qu'il enregistre et auxquels il réagit, il y a un ensemble spécifique de situations qui ont une influence beaucoup plus forte. Certains stimuli seraient entièrement responsables de la production, stabilisation et déclenchement d'une réaction. Par exemple, l'attitude difficile et antisociale d'un enfant peut être produite en « réponse à » et « renforcée par » l'attention des parents ou d'autres adultes significatifs⁴²⁰. C'est là bien un cas où on peut donner une signification psychologique claire à l'idée du « pouvoir de la situation » : le simple fait d'avoir perçu l'occurrence d'une situation déclenche des actions complexes et structurées qu'on aurait eu la tendance à expliquer par des facteurs psychologiques de plus haut niveau, tels que des intentions ou des stratégies de l'enfant. En ce sens, les effets de contrôle par le stimulus (qui sont très répandus et dont l'existence ne présuppose pas les prémisses problématiques du béhaviorisme) prouvent l'existence d'un côté clairement situationniste de notre psychologie. Mais est-ce que le situationnisme en psychologie sociale qui rend problématique la notion ordinaire de caractère est « situationniste » au sens du contrôle par le stimulus ? Pas du tout. Par la suite nous allons montrer que, paradoxalement, le situationnisme n'est tout simplement pas une thèse psychologique au sens de l'explication causale du comportement, il se limite à défendre une position situationniste sur la prévisibilité du comportement (affirmant notamment qu'on peut augmenter notre pouvoir de prévision en tenant compte des données sur ce que les gens ont fait dans des situations spécifiques, plutôt que de traits de caractère ou d'autres variables personnelles) qui a des conséquences normatives immédiates, tout en restant agnostique sur les causes véritables qui expliquent la conduite des gens. Or, ce n'est pas comme cela que la plupart des commentateurs voient le situationnisme (peut-être parce que ce n'est pas très flatteur). Pourtant, l'évidence textuelle est claire et les autres interprétations qui circulent sont soit incohérentes soit déconnectées des

⁴²⁰ Mischel, *Introduction to personality : toward an integration* 251.

expérience situationnistes (i.e. elles formulent des thèses cohérentes sur lesquelles les données ne peuvent pas trancher parce qu'elles ne sont pas pertinentes).

Interprétations causales du « pouvoir de la situation »

Commençons par les conceptions causales. La première interprétation qui a été retenue de « la force de la situation » est celle d'une explication causale par des facteurs physiquement extérieurs à l'individu, qui se trouvent littéralement « en dehors de sa peau ». L'idée serait que le comportement est expliqué tantôt par des états internes (croyances, désirs, intentions ou traits de caractères) tantôt par des facteurs externes. Je peux aller au cinéma parce que j'en avais envie, je l'avais planifié (internes) ou alors parce que le film est intéressant, un ami m'a demandé de l'accompagner, il pleut etc. (externes). Or, même s'il est vrai que dans nos explications ordinaires parfois on cite des états mentaux de l'agent sans faire référence à des facteurs externes et d'autres fois on fait l'inverse, il est clair que les deux styles explicatifs sont souvent interchangeable : si « je vais à une fête parce qu'il y a de la bonne musique » c'est aussi vrai que « je vais à une fête parce que j'aime la bonne musique » ou parce que « je désire écouter de la bonne musique ». En tout cas, quelles que soient les nuances pragmatiques (au autres) qui justifient le choix dans l'usage, on ne peut ni retenir ni attribuer au sens commun l'idée d'une explication causale (d'une action donnée) qui serait soit uniquement le fait de facteurs externes soit exclusivement le fait d'états mentaux internes, car l'idée est manifestement absurde : le comportement est toujours le résultat de l'interaction entre états mentaux et environnement⁴²¹. Dans cette interprétation l'opposition entre individu et situation est carrément incohérente. Même si dès 1977 Ross a critiqué la version causale de la dichotomie personne/situation⁴²², nous ne pouvons que constater avec surprise qu'une terminologie qui favorise cette confusion a été utilisée jusqu'en 1995 :

⁴²¹ L'idée selon laquelle l'un des deux facteurs pourrait être plus important que l'autre (i.e. c'est toujours la personne *et* la situation, mais parfois la situation (ou la personne) est plus décisive) dans des épisodes d'action déterminés est également inintelligible dans le cadre d'une explication causale. Cela n'a aucun sens de dire que les inputs externes sont plus importants que les processus mentaux internes, ou vice versa.

⁴²² Voir Lee D. Ross, Teresa M. Amabile et Julia L. Steinmetz, "Social Roles, Social-Control, and Biases in Social-Perception Processes," *Journal of Personality and Social Psychology* 35, no. 7 (1977). Ross critique Nisbett pour avoir pris au sérieux la valeur explicative de la distinction entre facteurs externes ou internes. Nisbett croyait avoir découvert une asymétrie entre la première et troisième personne : dans l'expérience, les participants expliquent leur choix d'une partenaire par des facteurs externes (il s'agit d'une femme très charmante) et le choix de leur meilleur ami par des facteurs internes (il aime les femmes très charmantes). Evidemment, cela ne signifie pas que – d'après les participants – deux types différents de causalité seraient concernés. Malheureusement, en théorie de

« Du côté ensoleillé de la peau il y a les forces externes ou situationnelles qui exercent une pression vers l'intérieur sur la personne, et dans le côté de la chair il y a les forces internes qui exercent une pression vers l'extérieur⁴²³. »

Il est également surprenant qu'en 2001 il soit encore utile de revenir sur la question⁴²⁴.

Une deuxième interprétation causale possible dont nous avons déjà anticipé l'échec est celle du contrôle par le stimulus. Le fait que des variations subtiles dans le contexte aient un effet disproportionné au niveau du comportement ne signifie pas que ce dernier soit renforcé ou déclenché par la perception de situations spécifiques. Dans l'expérience de Milgram, la variation de la distance de l'expérimentateur par rapport aux participants affecte le niveau d'obéissance (qui est plus élevé si l'expérimentateur est proche), mais cela ne signifie pas que le stimulus « un expérimentateur est très près de moi » explique à lui seul l'obéissance de la même façon que des signes d'attention de la part des parents sont suffisants pour expliquer et déclencher des comportements antisociaux complexes chez l'enfant. « Avoir remarqué qu'on est près d'un expérimentateur » n'a pas de propriétés psychologiques majeures et invariantes comme un stimulus de renforcement dans la théorie béhavioriste. Le fait qu'on ne fasse pas la même chose suivant qu'on est proche ou éloigné est trivial : en général on ne fait pas la même chose suivant qu'on est dans une situation ou dans une autre. D'après le béhaviorisme, la perception d'un stimulus est suffisante pour activer des comportements dont l'explication est normalement censée mobiliser des états ou processus mentaux complexes, tels que croyances, désirs ou traits de caractère. Dans les expériences situationnistes, il n'y a rien qui prouve que tels états et processus sont inactifs de la manière requise dans la théorie du contrôle par le stimulus.

Une variante possible de ce genre d'explication mérite d'être citée car elle a joué un rôle majeur dans l'histoire de l'étude psychologique du caractère. Il s'agit du modèle très modulaire d'explication du comportement humain développé par Thorndike, Terman et Woodworth au

l'attribution on a cru pendant des décennies que la différence entre explication situationnelle (externe) et explication dispositionnelle (interne) était fondamentale dans l'esprit des gens : une véritable distinction catégorique qu'ils appliqueraient à tout. Ainsi, une littérature énorme a étudié dans quelles conditions et par rapport à quels sujets on expliquerait plus ou moins en mobilisant des facteurs internes/externes. Conceptuellement, tout cela est incohérent et n'a pas de sens. Sur ce point voir B. F. Malle, "The actor-observer asymmetry in attribution: A (surprising) meta-analysis," *Psychological Bulletin* 132, no. 6 (2006).

⁴²³ D. T. Gilbert et P. S. Malone, "THE CORRESPONDENCE BIAS," *Psychological Bulletin* 117, no. 1 (1995).

⁴²⁴ Sabini, Siepmann et Stein, "The really fundamental attribution error in social psychological research."

tournant du siècle dans le cadre de leur critique à la psychologie des facultés⁴²⁵. Au niveau de la psychologie scientifique, vers la fin du XIX^{ème} siècle l'idée de traits de caractère « larges » et cohérents était étayée par la psychologie des facultés. Dans cette conception, le comportement humain serait expliqué par un nombre fini de facultés cognitives plutôt générales qui influenceraient de façon homogène, suivant leur niveau de maturation, une vaste classe de compétences. En pratique, cela signifie qu'une performance améliorée dans une tâche spécifique entraîne une amélioration générale dans d'autres domaines : si j'apprends le latin, je vais aussi être fort en grammaire française et en mathématique, car pour apprendre le latin il faut raisonner de manière logique et si cette faculté se développe toutes les performances relatives vont en bénéficier.

Au bout d'une série d'expériences au tournant du siècle Thorndike et Woodworth avaient montré que le transfert entre différentes tâches cognitives n'est que minimal et avaient développé une psychologie très modulaire: la cognition humaine ne serait qu'une somme de micro-capacités très spécifiques. Cela a été initialement appliqué aux fonctions cognitives relevant de l'intelligence générale, mais Terman, l'inventeur du concept de quotient d'intelligence, a soupçonné l'extension du phénomène et déjà parlé d'un « principe de fragmentation de la morale »⁴²⁶. Or, il s'agit là d'une véritable explication causale postulant, comme le feront les behavioristes, une architecture de la cognition humaine bien définie et informative. La littérature situationniste ne fait rien de tout cela, elle se limite à généraliser des effets comportementaux susceptibles d'être répliqués et compatibles, jusqu'à ce que la preuve du contraire, avec presque toute architecture cognitive.

Pour preuve, s'il en fallait, du fait que le lien entre des théories psychologiques et des programmes politiques et sociaux plus vastes est toujours moins évident et plus contingent qu'on ne le croit, ce « situationnisme » du début du siècle était fort réactionnaire. Le situationnisme d'aujourd'hui est franchement progressiste : il célèbre l'incohérence d'une évaluation générale qui partagerait les gens entre bons et mauvais dans l'absolu et célèbre la possibilité pour chacun de s'en sortir et de réussir, grâce à la manipulation des aspects spécifiques et sensibles des situations. Le situationnisme de Thorndike – un défenseur de l'eugénisme négatif – voyait plutôt dans la différence d'aptitude des gens dans des tâches spécifiques une raison... pour spécialiser les rôles dans la société! Cela devait passer par une spécialisation du curriculum scolaire : pourquoi enseigner le latin à tout le monde s'il y a des gens qui sont spécialement doués pour travailler dans des mines de charbon ?

⁴²⁵ Edward L. Thorndike et R. S. Woodworth, "The influence of improvement in one mental function upon the efficiency of other functions" *Psychological Review* 8 (1901). Lewis M. Terman, "Some paradoxes of personality, or muckraking in the psychology of character," *New England Monthly* 44 (1911).

⁴²⁶ Terman, "Some paradoxes of personality, or muckraking in the psychology of character," 374.

Autant leur apprendre des compétences de base, lire, écrire et un peu de mathématique élémentaire⁴²⁷.

Une troisième explication causale associe le pouvoir de la situation à l'incapacité d'agir sur la base d'engagements, désirs ou valeurs qu'on accepte et auxquels on s'identifie. En ce sens, on va dire que « c'est la situation » qui a fait que le sujet n'a pas respecté ses engagements réfléchis et ses valeurs. Dans l'expérience de Milgram, il est clair que la plupart des gens pensent qu'il ne faudrait pas aller jusqu'au bout avec les décharges électriques. Ceux qui suivent les ordres de l'expérimentateur apparaissent agités (parfois jusqu'à se déchirer les cheveux et rire de façon hystérique) et profitent de chaque occasion possible pour interrompre l'expérience (on constate que lors d'une interaction avec un expérimentateur physiquement plus éloigné, au téléphone, avec un expérimentateur qui n'a pas d'autorité suffisante ou avec deux expérimentateurs en désaccord entre eux le taux d'obéissance est réduit considérablement). Doris considère qu'un diagnostic de faiblesse de la volonté est adéquat dans le cas de plusieurs expériences :

« On a vu comment des facteurs situationnels non coercitifs peuvent faire en sorte que des personnes « ordinaires et décentes » agissent d'une façon qu'ils savent être injuste : les participants à l'expérience de Milgram se déchirent les cheveux pendant qu'ils électrocutent leurs victimes ; un geôlier dans l'expérience de la prison de Stanford était dégoûté de lui-même pendant qu'il abusait des prisonniers ; l'anxiété éprouvée par certains passants dans les expériences de Darley et collègues. Ces données suggèrent « faiblesse de la volonté », « incontinence » ou comme Aristote... l'a appelée, akrasia – des cas où une personne est consciente d'agir contre son meilleur jugement⁴²⁸. »

Cependant, il est douteux qu'on soit confronté ici à la véritable faiblesse de la volonté au sens classique, laquelle implique normalement que le sujet agisse à l'encontre d'un jugement spécifique dont il est conscient au moment de l'action (ex : je sais que je ne devrais manger de gâteau *maintenant* et je le fais quand même), ce qui n'est pas forcément le cas ici. Les participants éprouvent un conflit mais il n'est pas évident qu'ils se disent de façon claire « je ne devrais pas continuer ». Ils ont plutôt l'air d'être déchirés entre deux options⁴²⁹. Il s'agit d'un problème plus général d'autorégulation, c'est-à-dire la capacité à faire en sorte que nos valeurs et engagements

⁴²⁷ Voir Craig A. Cunningham, "A certain and reasoned art," in *Character psychology and character education*, édition Daniel K. Lapsley et F. Clark Power (Notre Dame, Indiana: University of Notre Dame Press, 2005), 179.

⁴²⁸ Doris, *Lack of character: personality and moral behavior* 134.

⁴²⁹ Dana K. Nelkin, "Freedom, Responsibility and the Challenge of Situationism," in *Free Will and Moral Responsibility* (Malden: Blackwell 2005), 195.

réfléchis soient suivis d'action concrètes (la faiblesse de la volonté est un défaut d'autorégulation spécifique et particulièrement grave).

Le problème de cette interprétation de la conclusion situationniste est représenté par les cas d'influence situationnelle moralement neutres ou positifs. Il arrive que des variations dans le contexte qu'on ne trouve pas significatives soient en mesure d'affecter le comportement des gens d'une manière positive ou simplement neutre : dans de tels cas on ne peut pas analyser « c'est la situation » par « cela va à l'encontre des engagements du sujet ». Les exemples d'influence positive sont des expériences où les gens ont cherché de rattraper un (faux, il collaborait à l'expérience) voleur après avoir pris un simple engagement informel de faire attention à un sac pendant quelques minutes (l'effet a été reproduit aussi à la plage et avec un voleur plutôt costaud), empêcher un viol, ou aider un technicien emmêlé dans des câbles électriques et qui apparemment venait d'être (faussement) électrocuté par accident. Dans ces trois cas l'attitude des participants est surprenante car ils prennent des risques et font quelque chose de très bien mais qui n'était pas requis : vraisemblablement la promesse de faire attention à un sac ne nous engage pas à risquer une lutte avec un voleur costaud (dont on ne sait même pas s'il est armé) et le devoir d'aider une personne en détresse ne nous engage pas non plus à détacher un individu de câbles électriques dangereux (le cas du viol est limite et on pourrait penser que l'intervention est requise : néanmoins, si on est tenu à chercher de l'aide il n'est pas clair qu'une intervention directe soit moralement requise).

Il y a aussi des cas d'influence situationnelle neutre où on est amené à faire quelque chose qu'on n'aurait pas fait sans l'influence de la situation, mais sans aller à l'encontre d'engagements forts ou clairement articulés : c'est le cas de la dissonance cognitive et des channeling factors (voir ci-dessus).

Une quatrième interprétation causale des expériences en psychologie sociale est critique par rapport à la conclusion que le comportement humain est contrôlé par l'influence de la situation, tout en soutenant que la littérature situationniste nous apprend des choses sur les processus causaux qui génèrent le comportement. D'après Sabini et ses collaborateurs ce n'est justement pas la situation, mais une série d'attitudes générales et sous-estimées des gens qui expliquent les résultats⁴³⁰. En particulier, l'impératif d'éviter l'embarras dans des situations sociales serait la cause de la plupart des effets situationnistes. Dans le paradigme expérimental de Milgram, le taux d'obéissance des participants est réduit si l'expérimentateur donne ses ordres par téléphone. Une explication plausible serait qu'il est très difficile de faire face et s'opposer à un expérimentateur présent

⁴³⁰ Sabini, Siepmann et Stein, "The really fundamental attribution error in social psychological research."

physiquement. On n'a pas le courage de refuser sa collaboration sans une justification valable à quelqu'un qui semble avoir l'autorité requise pour gérer la situation. Dans ce contexte, la difficulté à se justifier est accrue par la division en plusieurs étapes de la tâche : vu que les décharges électriques n'augmentent que progressivement, on a du mal à expliquer pourquoi on se refuserait de collaborer à un moment donné, plutôt qu'à l'étape précédente ou à l'étape suivante.

Dans les expériences de « conformité au groupe » de Asch ce genre d'explication est encore plus convaincant. Les participants apparaissaient très stressés et embarrassés lorsqu'il s'agit de contredire le groupe, aussi bien ceux qui parfois cédaient à la pression que ceux qui avaient le courage de donner la réponse correcte. Pour preuve de la difficulté à contredire le groupe, dans certaines répétitions de l'expérience, Asch avait fait donner une réponse incorrecte exprès par un collaborateur, pour constater la réaction du groupe (cette fois vrais participants) qui humilie le pauvre dissident en se moquant ouvertement de lui⁴³¹. Le soupçon qu'il s'agit probablement d'embarras plutôt que de véritable erreur est confirmé par les justifications données par les participants⁴³². L'embarras et la peur que le groupe se moque de nous sont aussi au cœur des expériences de dilution de la responsabilité dans le paradigme de Latané et Darley : si personne n'a l'air de remarquer la fumée qui sort d'une autre pièce, il y a probablement une raison, mais surtout on ne va certainement pas attirer l'attention de tout le monde pour signaler le problème, avec le risque que tout le monde se moque de notre naïveté.

Dans les expériences de dissonance cognitive les gens ont accepté pour un dollar d'écrire des essais où ils soutenaient des thèses contraires à leur position véritable, pour changer après d'idée en faveur d'une thèse plus en ligne avec l'opinion « forcée » de l'essai⁴³³. À l'interprétation situationniste qui voit les gens succomber à la pression situationnelle d'un dollar (et après changer d'idée à cause d'un phénomène de dissonance cognitive dû au fait de ne pas pouvoir reconnaître cette influence) Sabini préfère (à nouveau, c'est le même traitement du paradigme de Milgram) l'hypothèse que les gens se sentent embarrassés de refuser la collaboration avec l'expérimentateur : ce serait l'impossibilité à admettre cela qui déclenche la dissonance cognitive.

Bien sûr, on ne peut pas espérer pouvoir réinterpréter toutes les expériences situationnistes de cette façon, parfois l'influence de la situation n'a rien à voir avec l'embarras (si l'on n'est pas disposé à accepter une histoire trop tordue). Toutefois, cela ne constitue pas une objection profonde à la

⁴³¹ Asch, "STUDIES OF INDEPENDENCE AND CONFORMITY .1. A MINORITY OF ONE AGAINST A UNANIMOUS MAJORITY."

⁴³² Voir Hodges et Geyer, "A nonconformist account of the Asch experiments: Values, pragmatics, and moral dilemmas."

⁴³³ Sabini, Siepmann et Stein, "The really fundamental attribution error in social psychological research," 5.

stratégie de Sabini : le point est que la littérature situationniste ne prouverait en rien le pouvoir de la situation, elle montrerait tout simplement qu'une série de traits sont plus répandus et plus importants qu'on ne le croirait. Si les résultats ne s'expliquent pas uniquement par le rôle majeur de l'embarras (i.e. le désir d'éviter l'embarras) mais aussi par le désir d'être ponctuel, de faire plaisir aux autres (etc.) cela n'enlève rien à la critique, car alors le situationnisme se réduirait à la preuve de l'influence causale d'une liste de traits, sans aucune place véritable pour le concept de situation. Le problème de l'approche de Sabini est différent : ce qu'on entend par « influence de la situation » ne présuppose pas l'absence d'un thème de fond dans les « situations » qui l'exercent. Acceptons par absurdum que tous les cas typiques du situationnisme soient liés à l'embarras : cela n'autorise pas à postuler un trait de caractère du genre « tendance à éviter l'embarras », car il faudrait que ce trait rende prévisible le comportement des gens. Un trait qui nous permette de prédire les résultats expérimentaux et d'éviter des surprises situationnistes serait une raison suffisante pour considérer l'influence de la situation comme illusoire, mais si ces surprises persistent le fait qu'elles soient liées à un thème récurrent ne change pas grand chose. Une tendance à éviter l'embarras qui se manifeste d'une façon situationnellement imprévisible vient à l'appui du thème du « pouvoir de la situation ». La possibilité ou l'impossibilité de classer les influences situationnelles dans des groupes homogènes n'a tout simplement rien à voir avec ce qui fait problème dans les expériences, nous ne pouvons donc pas accepter la proposition de Sabini de voir la littérature situationniste comme la preuve du pouvoir causal d'une liste de traits qui auraient été ignorés.

Interprétations non causales : le rôle d'une référence aux situations dans la prévision du comportement

Une interprétation non-causale répandue⁴³⁴ se base sur la fréquence des actions dans une situation donnée. Si l'on y réfléchit, en disant que c'est à cause de la situation qu'on accepte de suivre des ordres avec un revolver braqué sur la nuque l'on veut probablement dire que tout le monde ferait la même chose dans cette situation. On peut alors adopter la définition suivante : le comportement est causé par la situation s'il s'agit de ce que la plupart des gens feraient (dans des circonstances données), sinon il est le fait des dispositions de la personne.

⁴³⁴ On la retrouve en E. E. Jones et K. E. Davis, "FROM ACTS TO DISPOSITIONS - THE ATTRIBUTION PROCESS IN PERSON PERCEPTION," *Advances in Experimental Social Psychology* 2, no. 4 (1965), Ross, Amabile et Steinmetz, "Social Roles, Social-Control, and Biases in Social-Perception Processes."

Cette définition semble expliquer ce que nous voulons dire par « c'est la situation » dans un bon nombre de cas : il serait tentant de supposer que les psychologues sociaux veulent dire la même chose par rapport aux effets expérimentaux produits dans la littérature situationniste. Malheureusement, il est très facile de trouver des contre-exemples. Dans cette conception le point crucial serait qu'il y a beaucoup plus d'uniformités dans le comportement qu'on ne le croit ou, mieux, qu'on cherche l'uniformité au mauvais endroit. Ainsi, on pense que les gens sont différents et que cette différence pourra se manifester dans chaque circonstance. Mais il se trouve que dans beaucoup de cas dans lesquels on s'attend à ce que les qualités ou les vices ressortent de manière significative, la plupart des gens font la même chose. Ignorer la « force des situations » sera donc surestimer la variabilité du comportement dans toute une série de circonstances cruciales. Or, dans plusieurs des expériences les plus célèbres du situationnisme, il arrive au contraire que les gens *sous-estiment* la variabilité réelle. Par exemple, dans l'expérience de Milgram les gens prévoient que seuls 2% ou 3% des participants vont continuer jusqu'à la décharge électrique maximale, alors qu'environ deux tiers obéissent jusqu'au bout. On retrouve le même schéma dans les études de Asch sur l'influence sociale dans le jugement et de Latané et Darley sur la dilution de la responsabilité : les gens s'attendent à ce que pratiquement tout le monde fasse la même chose, notamment résister à la pression du groupe, alors que le comportement est plus varié que cela : il y en a qui cèdent et il y en a qui résistent. Cependant, il s'agit de cas paradigmatiques d'influence de la situation, influence qui donc n'est pas caractérisée correctement par des faits relatifs à la fréquence et à la variation individuelle⁴³⁵.

Une autre version déguisée du thème de la variabilité du comportement est l'idée que « c'est la situation » peut être paraphrasé par « ce qui c'est passé ne nous donne pas de raisons de postuler des différences individuelles ». Avec un revolver braqué sur la nuque, ce qu'on fait est expliqué par la situation parce qu'on n'est pas autorisé à tirer des conclusions sur nous à partir de notre comportement dans ces conditions. En anglais, on dirait que « c'est la situation » quand le comportement n'est pas *individuating* (il n'est pas pertinent pour cerner la spécificité d'un individu). Cette interprétation ne marche que dans les situations où tout le monde ferait la même chose, mais nous avons vu que ce n'est pas toujours le cas dans les expériences situationnistes, où le pourcentage de gens qui font ou ne font pas une chose n'est proche ni de 0 ni de 100.

⁴³⁵ On pourrait vouloir maintenir que les gens commettent bien une erreur relative à la variabilité du comportement : ils ne se rendent pas compte du fait que deux tiers des participants vont jusqu'au bout dans l'expérience de Milgram. Mais il s'agit d'une analyse tautologique : une erreur de prédiction est par définition relative à la variabilité du comportement.

Venons-en finalement à l'interprétation des résultats situationnistes que nous considérons tenable. Dans une expérience situationniste typique ce qui se passe est que nous sommes surpris par ce que les gens font. Les attentes (aussi bien du sens commun que des expérimentateurs, si les deux diffèrent) sont déçues : quelle que soit notre théorie populaire du comportement et les heuristiques cognitives que nous déployons pour anticiper les actions d'autrui, elles nous donnent un mauvais résultat. Cependant, grâce au fait d'avoir conduit l'expérience les psychologues découvrent que, si on utilise une stratégie de prédiction qui fait référence directement à la variation du comportement dans la situation expérimentale, on peut améliorer la performance de prédiction.

Chaque expérience montre qu'il y a une situation x telle que :

1. nos stratégies cognitives sont mauvaises pour prédire ce que les gens feraient dans la situation x
2. une stratégie de prédiction qui utiliserait directement les données sur la variation du comportement en x serait meilleure

Pourquoi appeler cela situationnisme ? Le point essentiel est que, pour l'instant, la référence à une situation spécifique x (par rapport à une expérience, ou une liste de situations $x, y, z \dots$ plus généralement) est *non-éliminable*. En négatif, il n'y a pas d'autres façons de caractériser le comportement des gens que de dire qu'ils se montrent imprévisibles dans la situation x . En positif, il n'y a aucun conseil à donner sinon d'utiliser directement ce qu'on vient de découvrir sur la situation x . Si on pouvait éliminer la référence à une liste de situations spécifiques (par exemple, en soutenant comme Sabini que les gens ne veulent pas être trop embarrassés et qu'il faudrait tenir compte de cela pour prédire le comportement) le situationnisme disparaîtrait.

Ce qui ressort de la littérature psychologique est bien un situationnisme digne de ce nom, mais la notion informative de situation concernée n'a rien à voir ni avec une ontologie de situations objectives qui existeraient dans le monde (particularistes moraux) ni avec un pouvoir *causal* psychologique de stimuli situationnels. Il s'agit simplement du rôle qu'une référence explicite à une liste donnée de situations joue dans des questions de prévisibilité du comportement humain. Même si elle n'est pas une thèse psychologique au sens fort (elle ne décrit pas des processus causaux réels), l'imprévisibilité situationnelle a des conséquences psychologiques indirectes : simplement, toute théorie psychologique qui fait des prévisions incompatibles ne peut être adéquate. Il s'agit toutefois d'un critère de correction totalement négatif : la littérature situationniste ne nous dit rien sur les causes psychologiques de l'imprévisibilité chez les sujets dont le comportement résulte imprévisible. Nous sommes maintenant en mesure de comprendre que la façon standard de présenter le situationnisme comme thèse substantielle sur le pouvoir de la situation est trompeuse : il n'y a pas de situations tout-puissantes qui nous contraignent et auxquelles on peut éventuellement

résister, il y a seulement des situations où notre pouvoir de prédiction nous trompe sans qu'on sache pourquoi (au moins, si on se limite à ce qu'on peut apprendre de la littérature situationniste).

Le fait de ne pas impliquer une ontologie des situations donne aussi un avantage au situationnisme en psychologie sociale par rapport à d'autres versions. En particulier, il évite le cauchemar de toute théorie qui donne un rôle majeur au contexte : expliquer ce qu'est une situation. Gérer le vague de la notion ordinaire de situation pour arriver à une théorie ontologique ou psychologique informative sur le contexte a toujours été difficile. Pour comprendre le problème, il suffit de penser à des affirmations du genre : « la signification d'une phrase dans un livre dépend du contexte ». Quel est le contexte pertinent ? Le paragraphe, la page, le chapitre, ou tantôt l'un et tantôt l'autre ? Il semble qu'une définition intéressante de situation souffre d'un problème de variabilité contextuelle de deuxième degré : le genre de facteurs qui comptent comme pertinents dépendent aussi du contexte. Le situationnisme évite ce problème car le format de la situation pertinente est donné simplement par la description de la situation qui permet de reproduire l'effet de manière fiable. Si on découvre un effet situationnel intéressant « à la plage lorsqu'on vous demande une bière » cette situation verra s'ajouter à la liste des situationnistes, s'il est nécessaire de spécifier dans les détails le paradigme de Milgram pour reproduire l'effet, alors sera cette description complète qui entre dans la liste etc. Il ne doit pas y avoir un format fixe, il s'agit juste des contextes où le comportement se montre imprévisible.

Nonobstant le manque de profondeur psychologique (tant qu'une meilleure compréhension des causes n'émerge pas), les ambitions normatives du situationnisme sont claires : d'un point de vue normatif, il paraît évident qu'une stratégie de prédiction plus performante doit être préférée à une stratégie inefficace. Si les expériences prouvent qu'environ deux tiers des sujets dans le paradigme de Milgram obéissent, il s'ensuit que les gens devraient intégrer cette donnée dans le cadre de leur stratégie de prédiction. Ils devraient apprendre que dans des circonstances « comme cela » la plupart des gens acceptent de collaborer, non pas simplement que « les gens sont plus soumis à l'autorité qu'on ne le croit ». Même si ce dernier constat était révisionniste par rapport à l'attitude des gens, il n'aurait rien de situationniste : la subtilité réside dans le fait que la référence à une liste de situations est cruciale.

On peut maintenant définir avec précision le situationnisme en tant que thèse spécifique sur laquelle convergerait toute la littérature situationniste :

Situationnisme = imprévisibilité situationnelle généralisée : « Le comportement humain et les stratégies actuellement disponibles pour le prédire sont tels qu'il existe une liste indéfiniment longue et indéfiniment extensible de situations dans lesquelles ce que les gens feraient en moyenne est imprévisible »

III. Le situationnisme contre la théorie de la vertu

Après avoir compris le situationnisme en tant que thèse et son lien avec la littérature situationniste au sens large, il faut maintenant analyser pourquoi le projet d'une théorie des vertus serait mis à mal par les données situationnistes. La situation est paradoxale parce que si d'un côté tout le monde voit intuitivement dans cette littérature une menace énorme pour la théorie de la vertu, de l'autre il est très difficile de mettre en évidence avec précision le problème et d'explicitier les arguments qui conduiraient des expérience situationnistes au rejet de la théorie des vertus. A ce sujet il y a eu beaucoup de confusion dans les réactions des défenseurs d'un projet de construction du caractère, confusion dont les situationnistes jouent abondamment, sachant qu'un manque de précision dans la formalisation de leur critique ne peut que les favoriser. La réaction intuitive à ces données jouant à leur faveur, une précision excessive ne peut que gaspiller ce capital de plausibilité en offrant des possibilités de falsification faciles.

Ainsi Doris prend soin de refuser explicitement toute formulation trop claire de sa stratégie : il ne s'agit pas d'attaquer la théorie de la vertu au nom du principe « à l'impossible nul n'est tenu », car aucune preuve empirique ne pourra montrer l'impossibilité stricte de développer des traits de caractères. On apprend qu'il ne s'agit pas non plus de mobiliser une version plus atténuée du même principe :

« Autrement, « à l'impossible nul n'est tenu » peut être compris comme une prohibition contre des prétentions éthiques non raisonnables ou injustes, et paraphrasé plus ou moins de la façon suivante : « pour qu'une obligation soit valide, cela doit être raisonnablement possible de la suivre ». Mais, pour appliquer ce principe, il faut trancher au cas par cas sur le caractère raisonnable d'une obligation. Le défenseur de l'éthique de la vertu ne se trouverait pas dans une position difficile : si des idéaux inatteignables de vertu inspirent les gens à se comporter mieux, il y a une bonne possibilité de montrer qu'ils sont raisonnables⁴³⁶. »

En positif, Doris semble avoir à l'esprit l'idée d'un équilibre réfléchi « large » dans lequel on ferait entrer aussi bien les arguments philosophiques classiques de l'éthique de la vertu que les données psychologiques :

« Mon approche évoque la méthode de l'équilibre réfléchi de Rawls, où la réflexion éthique recherche un équilibre des considérations générales tirées de la théorie éthique, des

⁴³⁶ Doris, *Lack of character: personality and moral behavior* 113.

*jugements éthiques spécifiques, des meilleures théories de la psychologie individuelle, des dynamiques sociales et des circonstances matérielles*⁴³⁷. »

Concrètement, cela veut dire que si l'on prend en considération en même temps les aspirations de la théorie des vertus et les données empiriques situationnistes en cherchant à arriver à un cadre cohérent (pour arriver à ce cadre, toute croyance qui entre dans le processus d'équilibre réfléchi peut être abandonnée), on va perdre toute motivation de conserver les aspirations traditionnelles.

Or, ce n'est pas parce personne ne demande un argument déductif formalisé (auquel la complexité d'un équilibre global de la cohérence de nos croyances ne peut se réduire) – et que tout le monde voit intuitivement la menace- qu'on peut accepter de s'en tenir. Il n'y aura pas d'argument décisif unique, mais un processus d'équilibre réfléchi n'est ni magique ni insondable : il faudra bien pouvoir expliquer quel genre de considérations militent contre la théorie de la vertu. Il me semble que les situationnistes ont abusé du droit à ne pas formaliser un ou plusieurs « arguments décisifs » pour ne pas être aussi clairs qu'ils auraient pu, alimentant ainsi des confusions majeures dans le débat.

Nous allons défendre l'idée selon laquelle le seul argument puissant contre la théorie des vertus qu'on peut construire à partir des données situationnistes passe par une interprétation radicale de la thèse situationniste. En d'autres termes, la vraie menace vient uniquement d'une interprétation radicale de l'imprévisibilité situationnelle du comportement humain : tout autre argument qu'on peut envisager à partir d'une partie des données - isolées de leur soutien à l'imprévisibilité situationnelle *radicale* - n'est pas convaincant. En ce sens, les doutes situationnistes sur le caractère moral sont la conséquence d'une propriété générale du comportement humain. Nous allons montrer que, si notre psychologie n'était pas radicalement situationniste en général, les données relatives à une variabilité situationnelle spécifique au comportement moral ne poseraient pas de problèmes majeurs pour une théorie des vertus. Réciproquement, une fois qu'on a admis le situationnisme radical en général, ces mêmes données n'ajoutent non plus grand chose à la compréhension de la question. Le débat se joue entièrement au niveau de la validité d'une conception générale de la psychologie humaine.

La fragmentation radicale du caractère

Le situationnisme est l'idée selon laquelle les phénomènes d'imprévisibilité situationnelle mis en évidence dans la littérature situationniste sont très généraux : s'il s'agissait de dire qu'ils se limitent

⁴³⁷ Ibid. 114.

aux cas étudiés expérimentalement ou à une série circonscrite de situations semblables il n'y aurait pas même de thèse intéressante ou provocatrice à discuter. Mais quel genre et quel niveau de généralité ont-ils ? Il se trouve que répondre à cette question est vital pour comprendre la portée de la menace situationniste. La première étape à franchir est de prouver la validité écologique des expériences. La quasi-totalité de l'attention des critiques se concentre sur ce point, le genre de question étant par exemple « est-ce qu'il y a des signes d'obéissance à la Milgram dans le monde réel ? », « dans quelle mesure les gens vont-ils céder à la pression du groupe dans des contextes moins formels qu'une expérience de laboratoire ? ». Normalement, être rassuré sur la validité écologique est tout ce qu'il faut à des processus psychologiques qui ont fait leurs preuves dans le laboratoire pour conquérir le monde réel : si une découverte intéressante liée à la cognition de la lecture ne montre pas de problèmes de validité écologique, c'est que les processus concernés sont activés lorsque les gens lisent dans la vie de tous les jours. Cependant, dans le cas du situationnisme il y a une complication. Le tissu de la vie quotidienne des individus et d'une société est composé par une infinité de situations, et le situationnisme est une thèse portant simplement sur l'existence d'une série indéterminée de situations d'un certain genre, où se vérifient des phénomènes spécifiques (imprévisibilité situationnelle). On peut bien admettre que des situations à la Milgram existent dans la vie réelle : le situationnisme a besoin de bien plus que de la validité écologique quant à l'*existence* d'une classe de phénomènes, il doit en prouver la représentativité. Pour utiliser une métaphore spatiale, on peut imaginer le tissu de toutes les situations qu'on peut rencontrer dans la vie quotidienne comme un espace tridimensionnel et les situations où l'imprévisibilité situationnelle se manifeste comme des particules. Pour être une thèse provocatrice le situationnisme a besoin de soutenir que les particules sont distribuées et saturent de façon plus ou moins homogène l'espace des situations.

Dans cette métaphore, l'homogénéité de la distribution est plus importante que la densité, car le situationnisme ne s'engage pas à soutenir qu'on doit faire face à l'imprévisibilité littéralement toutes les dix minutes, l'idée étant plutôt qu'elle menace constamment la réalisation de tout projet et activité structurés. Comme nous l'avons vu par rapport à la critique de Sabini à la littérature situationniste, le situationnisme pourrait être cohérent et informatif même s'il y avait un nombre limité de thèmes communs aux situations où l'imprévisibilité se manifeste (ce qui n'est quasi certainement pas le cas), par exemple si des manifestations d'embarras étaient toujours à la base de l'imprévisibilité situationnelle. Même dans ce cas de figure, encore faudrait-il que les circonstances où l'embarras nous joue un mauvais tour soient distribuées de façon suffisamment homogène dans l'espace des situations : il faut que, dans la vie de tous les jours, il y ait toujours un danger et une possibilité constante de tomber dans l'une de ces circonstances. Sans cela, on voit mal l'intérêt de

situationnisme et on pourrait rétorquer qu'il existe une infinité d'ordre supérieur de situations où l'on ne céderait pas à des influences situationnelles : par exemple, il n'est pas facile de faire changer d'intention de vote à un militant ou de convaincre quelqu'un d'affecter son salaire à des œuvres de caritatives par des pressions triviales. Dans une lecture minimale, le situationnisme ne fait qu'une affirmation relative : « il y a beaucoup plus d'imprévisibilité situationnelle que l'on ne *croit* ». Si on s'arrête là, sans aucune idée sur l'ampleur de phénomène dans une échelle absolue, on n'en dérive rien : il est possible que nos attentes minimiseraient de façon exagérée l'imprévisibilité existante mais que la valeur réelle reste compatible avec un projet de construction du caractère.

C'est cette idée de saturation de l'espace des situations qui permet de passer du situationnisme à la thèse de la fragmentation du caractère. Il n'est pas possible qu'il y ait de traits de caractère larges avec une grande cohérence inter-situationnelle si l'on finit par tomber à chaque fois dans des pièges situationnels : la cohérence du comportement des gens est menacée et limitée à tous azimuts. Il ne s'agit pas de dire qu'il y a beaucoup plus de pollution que l'on ne croyait, mais que dans l'absolu le monde est trop pollué pour que la vie continue comme avant. S'il n'y avait pas d'homogénéité de distribution, il y aurait des secteurs de la vie où développer des traits larges serait possible et souhaitable : imaginons par absurdum que l'imprévisibilité (avec ou sans thèmes de fond à la Sabini) soit concentrée uniquement dans l'exercice de certains rôles, par exemple que les gens seraient particulièrement sensibles à la pression de la situation en tant que patrons d'une entreprise. Dans ce cas, il serait possible de contourner la menace et développer des traits de caractère très larges dans tous les autres domaines : si la cohérence inter-situationnelle de ces traits ne peut s'étendre à la gestion d'une entreprise – qui est soumise au pouvoir de la situation – il ne s'agira pas de traits identiques à ceux du sens commun, mais développer une honnêteté « sauf-entreprise » est évidemment un projet intéressant si on ne peut avoir mieux. D'ailleurs, les conceptions de l'honnêteté défendues dans la tradition de la théorie de la vertu ne diffèrent pas plus entre elles et avec le sens commun que celle-ci.

Si une lecture du situationnisme qui n'irait pas même dans le sens d'une diffusion plutôt homogène de l'imprévisibilité serait un *no-starter*, pour avoir un bon argument contre la théorie des vertus il faut concevoir l'espace de l'action morale comme étant bien saturé de menaces situationnistes. Le situationnisme ne peut pas prouver qu'il n'y a pas de façon de résoudre le problème posé par une liste limitée de situations ou de secteurs de la vie bien définis où serait confinée l'imprévisibilité, avec une poignée de situations qui perturbent parfois le reste de la vie morale. Un projet d'éducation du caractère peut mobiliser des ressources considérables et la difficulté de la tâche peut stimuler et donner une impulsion ultérieure pour affronter les épreuves d'un chemin vers la maturité morale.

Le point n'est donc pas la difficulté du processus en tant que telle, mais plutôt la difficulté d'un processus qui ne fait pas de sens. Si la lutte contre la variabilité situationnelle était constructive, avec la possibilité de progresser, de perdre des batailles, mais aussi de recommencer et de progresser en construisant des zones de plus en plus étendues de cohérence et autonomie dans sa propre personnalité, il faudrait sans doute s'y mettre. Mais si l'espace moral est saturé d'imprévisibilité situationnelle, cela signifie que le combat est perdu d'avance, la tentative de construire de la cohérence y serait continuellement frustrée. Je cherche à être un ami honnête, et ensuite à vivre mon honnêteté également en tant que citoyen. Non seulement l'extension du champ de ma vertu n'est pas garantie malgré l'effort, le premier résultat n'est jamais consolidé non plus, mon honnêteté en tant qu'ami n'étant pas à l'abri de facteurs situationnels qu'on ne peut contourner d'aucune façon systématique. Dans ces conditions la construction d'un caractère moral rencontre un genre de difficultés qui rendent le projet dénué de sens. Ce constat ouvre la voie à l'éthique situationniste radicale défendue par Doris, Harman et les réalistes critiques en théorie du droit⁴³⁸.

Dans une interprétation correcte, l'argument situationniste part de la tendance générale dans la littérature, passe nécessairement par une lecture radicale de l'imprévisibilité situationnelle et en dérive une image radicalement fragmentée du caractère, dont la cohérence inter-situationnelle ne peut pas être reconstruite. Plusieurs auteurs ont donné des interprétations différentes de l'argument censé relier la littérature situationniste au rejet de la théorie de la vertu. Nous allons maintenant les analyser et les critiquer.

Des traits de caractère conçus de façon « behavioriste » n'existent pas

Annas, Webber et Kamtekar⁴³⁹ croient que l'argument situationniste vise uniquement l'existence de trait de caractères définis à la manière behavioriste comme « *dispositions à agir de façon stéréotypée et qui fonctionnent indépendamment l'une de l'autre*⁴⁴⁰. » Ils sont convaincus de cela à cause de la structure typique d'une expérience situationniste : il y a toujours une action normativement requise que le sujet va ou ne va pas faire (refuser de continuer dans l'expérience de

⁴³⁸ Ibid, Jon Hanson et David Yosifon, "The Situation: An Introduction to The Situational Character, Critical Realism, Power Economics, and Deep Capture," in *Santa Clara University School of Law Legal Studies Research Papers Series* (2006), Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99 (1999).

⁴³⁹ Annas, "Virtue ethics and social psychology." Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character." Webber, "Virtue, Character and Situation."

⁴⁴⁰ Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character," 460.

Milgram, dire la vérité contre l'opinion du groupe dans le paradigme de Asch). Du fait que beaucoup de sujets ne font pas ce à quoi on s'attend, les situationnistes semblent déduire que les traits de caractère visés ne sont pas très répandus, il paraît donc que la notion de trait qu'ils retiennent est celle de disposition mécanique à faire certaines actions dans des contextes donnés.

Cette attribution est fallacieuse. Les situationnistes ne présupposent aucune conception spécifique des traits, ils s'attendent uniquement à ce que la possession pleine d'un trait large laisse des signes reconnaissables au niveau du comportement extérieur. Doris écrit :

« ...décrire les vertus comme des dispositions au comportement c'est en donner une image très déformée ; on admet généralement que la vertu a affaire non seulement à ce qui se passe « à l'extérieur » sous la forme de comportement explicite mais aussi à ce qui se produit « à l'intérieur » sous la forme de motifs, émotions et cognitions »

et « Les théories qui mettent l'accent sur des processus « internes » doivent être comprises comme compléments des théories qui soulignent les dispositions au comportement⁴⁴¹. » L'idée est que l'action testée dans l'expérience doit être diagnostique du trait de caractère visé, mais pas parce que les traits seraient des dispositions mécaniques. On part du présupposé qu'être honnête doit impliquer au moins et entre autres qu'on ne va pas tout faire pour voler un ami. Le comportement extérieur qui est prima facie pertinent pour l'attribution d'un trait n'est pas toujours très central ou diagnostique : on s'attend plutôt à ce que quelqu'un de sociable discute tout le temps et avec tout le monde à une fête, mais est-ce que cela veut dire qu'on va douter de sa sociabilité si pour une fois il se tait pendant une demi-heure ou plus ? Non. Il y a plusieurs explications possibles pour un épisode de ce genre qui sont compatibles avec sa possession du trait de caractère concerné. Sans davantage de précisions, tel que nous l'avons caractérisé, l'épisode n'est pas révélateur, et si quelqu'un le considère comme significatif c'est peut-être parce qu'il pense que posséder un trait signifie agir mécaniquement d'une certaine façon dans les circonstances pertinentes. Mais les situationnistes présupposent uniquement que quelques situations fortement diagnostiques doivent bien exister et on doit pouvoir en donner des exemples. Nier cela revient à penser que la vraie et complète possession d'un trait tel que l'honnêteté n'assure rien de précis au niveau du comportement extérieur facilement vérifiable, ce qui est absurde : on pourrait concocter une histoire où un vrai honnête vole de l'argent à son meilleur ami pour des raisons futiles.

Par contre, une objection acceptable consiste à critiquer au cas par cas le choix des situations fortement diagnostiques pour un trait donné dans les expériences. Cette voie a été parcourue avec un certain succès pour les expériences qui concernent le comportement moral. Adams remarque de

⁴⁴¹ Doris, *Lack of character: personality and moral behavior* 16-17.

façon convaincante qu'aider ou ne pas aider quelqu'un qui laisse tomber des papiers dans la rue n'est tout simplement pas suffisamment significatif et diagnostique d'un point de vue moral⁴⁴². Non seulement parce que ce n'est pas très grave, mais parce que aider autrui est une « obligation imparfaite », c'est à dire un devoir qui nous laisse la discrétion sur comment mieux agir, alors que dans le cas d'une obligation parfaite une certaine action est requise. Respecter des obligations imparfaites laisse la place pour des stratégies qui sont compatibles avec un comportement superficiellement non-moral et non-coopératif dans des cas particuliers, par exemple on peut choisir de faire l'aumône avec certaines catégories de gens dans le besoin et pas avec d'autres, ou ne pas faire l'aumône du tout mais aider d'autres façons. Il n'est pas clair qu'on ait une obligation parfaite d'aider à ramasser des papiers, et il est au moins discutable qu'on ait une obligation parfaite de s'arrêter dans le paradigme du « Bon Samaritain ». Dans ce dernier cas, il était prévu que la victime apparaisse « ...*ambiguë – mal habillée, probablement ayant besoin d'aide, mais aussi peut-être ivre ou même potentiellement dangereuse*⁴⁴³. » Il n'est pas certain qu'on soit moralement obligé à aider quelqu'un qui paraît dangereux, en tout cas il est légitime de douter du caractère diagnostique de cette situation.

Le paradigme de Milgram paraît un cas de situation diagnostique par excellence, vu que la plupart des gens s'accordent sur le fait que collaborer est moralement inacceptable et que la situation n'excuse pas. En ce sens, il s'agit du chef d'œuvre de la tradition situationniste. La difficulté de prouver qu'une expérience donnée est vraiment diagnostique de l'absence d'un trait spécifique nous donne une raison de plus pour penser que la stratégie situationniste doit se baser sur le constat général de l'imprévisibilité situationnelle de tout comportement humain. En d'autres termes, s'il s'agissait de viser l'un après l'autre le courage, l'honnêteté, la franchise en exposant les sujets à des situations diagnostiques de ces vertus, l'argument situationniste serait faible. Les expériences sont très biaisées car elles se concentrent essentiellement sur la compassion (ou en tout cas, de différentes formes de l'attitude à aider autrui) et même dans ce cas souvent, comme nous l'avons vu, l'aspect diagnostique est contestable. Le vrai point est que de vraies situations diagnostiques où les sujets échouent doivent forcément exister et devraient pouvoir être repérées, *si* le comportement humain est situationnellement imprévisible, comme la littérature situationniste dans son ensemble le montre.

⁴⁴² Adams, *A theory of virtue : excellence in being for the good* 145-148.

⁴⁴³ Darley et Batson, "FROM JERUSALEM TO JERICHO - STUDY OF SITUATIONAL AND DISPOSITIONAL VARIABLES IN HELPING BEHAVIOR," 102.

Les traits de caractère n'existent pas : l'incompatibilité entre l'existence du caractère et de la vertu et les résultats situationnistes qui concernent le comportement moral

La façon la plus répandue d'exposer l'argument situationniste consiste à dire que les expériences traitées par Doris et Harman prouvent la non-existence des traits de caractère : comme les vertus sont des traits de caractères, une théorie des vertus est vouée à l'échec. Ainsi formulé, l'argument est clairement non pertinent. La théorie de la vertu s'intéresse à la possibilité de construire et de développer des traits de caractères, elle ne présuppose pas l'existence préalable de traits sur lesquels intervenir. La condition préalable pour pouvoir devenir honnête n'est pas d'être vicieux et uniformément malhonnête, plutôt que simplement instable, immature et incohérent dans son attitude. La conclusion d'une attaque contre la théorie de la vertu doit porter sur la possibilité de *développer* des traits de caractère, non pas sur leur existence factuelle dans la population. Malgré cette confusion, l'argument peut être amendé en postulant une prémisse implicite : s'il n'y a pas de traits de caractère, ou s'ils sont très rares, il doit y avoir une raison et cela ne va pas être très facile d'en développer. Il est évident que c'est bien cela que les situationnistes ont à l'esprit. Cette prémisse implicite est-elle plausible ? Au premier abord, il ne va pas de soi et paraît même très pessimiste de supposer qu'on ne puisse pas développer des structures de caractère qui ne sont pas répandues. Cependant, dans le présent contexte nous allons concéder cette prémisse, admettant ainsi que si les traits sont absents ou très rares, on a un bon argument contre la théorie de la vertu. Car lorsqu'on parle d'excellence morale on se réfère aussi et surtout à un phénomène qui existe aujourd'hui et qui a bien existé dans l'histoire. On veut rendre compte des cas paradigmatiques et réels d'excellence dont on fait l'expérience et on a le témoignage. Il peut être légitime d'imaginer une dimension surérogatoire et supérieure de moralité qu'on pourrait atteindre dans l'avenir et qui ne trouve pas d'exemplification aujourd'hui, mais rendre compte de l'excellence morale telle qu'elle existe est une contrainte fondamentale de toute théorie éthique actuelle. S'il n'y a pas ou très peu de traits, c'est qu'ils n'ont pas été développés (qu'il soit potentiellement facile ou difficile de le faire) et donc ils ne peuvent pas expliquer le phénomène de l'excellence morale, dont on ne peut dire qu'il serait presque inexistant.

Athanassoulis⁴⁴⁴ a interprété l'argument situationniste comme affirmant l'incompatibilité directe entre ce qui se passe dans les expériences situationnistes (non en général, celles relatives à la cognition morale) et l'existence du caractère et de la vertu. Comme il arrive qu'une expérience scientifique ne puisse pas expliquer certains résultats, l'hypothèse de l'existence du caractère ne pourrait pas expliquer le comportement dans le paradigme de Milgram ou de Latané. Toute chose égale par ailleurs, elle doit être rejetée. Il n'y a donc pas de traits de caractère.

Cette interprétation est à attribuer en partie à Harman, qui considère que le pourcentage d'obéissants dans l'expérience de Milgram (tout comme celui d'autres comportements condamnables dans les expériences situationnistes) est trop élevé pour qu'on puisse attribuer l'obéissance à un défaut de caractère⁴⁴⁵. Il a l'air de vouloir dire que ce résultat est incompatible avec une explication caractériologique, c'est le genre de chose qui ne peut arriver si les gens ont un caractère. Cela est trivialement faux, l'obéissance pourrait être expliquée par toute sorte de conflit interne entre des vertus qui ne sont pas encore pleinement développées (compassion, respect pour autrui) et des traits larges (le respect pour l'autorité). En d'autres termes, il est possible d'expliquer l'attitude des gens par l'existence d'*autres* traits que ceux dont la cohérence est directement visée.

Il est en général très facile de réinterpréter une expérience situationniste en termes d'activation et de conflit entre des traits de caractère. Tout ce que l'expérience montre est qu'une explication déterminée qui invoque certaines vertus pleines ne peut pas être correcte : notamment, il n'est pas possible que la plupart des participants possèdent la vertu de la compassion. Toutefois, ce qu'ils font peut très bien être expliqué par la possession d'autres traits larges en conflit avec des tendances vertueuses qui ne sont pas encore mûres et en mesure de s'imposer. Le problème de la stratégie de description des expériences situationnistes en termes d'autres traits (que ceux dont l'existence est directement visée) n'émerge que si l'on mobilise la thèse de l'imprévisibilité situationnelle radicale. Si l'espace des situations est saturé par des pièges situationnels c'est que ce deuxième groupe de traits qu'on a postulé ne peuvent non plus exister, car on pourrait toujours reproduire des expériences qui les visent et montrer ainsi leur fragilité face à la force de la situation. Mais si la littérature situationniste n'est pas mobilisée dans sa totalité (non seulement la partie qui concerne le comportement moral) pour étayer la thèse de l'imprévisibilité radicale, la possibilité d'une réinterprétation ne peut être définitivement écartée. L'argument des situationnistes n'est pas que

⁴⁴⁴ Nafsika Athanassoulis, "A Response to Harman: Virtue Ethics and Character Traits," *Proceedings of the Aristotelian Society* (2000).

⁴⁴⁵ Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error."

l'existence du caractère est incompatible avec une liste déterminée de résultats expérimentaux relatifs au comportement moral⁴⁴⁶.

Les traits de caractères moraux sont extrêmement rares

Une autre interprétation très répandue veut que les expériences citées par Doris et Harman (toujours par rapport au comportement moral, non pas la littérature situationniste dans son intégralité) seraient directement incompatibles non pas avec l'existence mais avec la diffusion des traits : si des phénomènes de ce type se produisent, c'est qu'il y a forcément très peu de gens qui ont des traits de caractère. Pour une critique de la théorie des vertus la rareté extrême aurait ensuite, comme nous l'avons dit, les mêmes conséquences que l'absence totale. Qu'il y en ait peu ou pas du tout, les traits de caractère ne peuvent pas rendre compte de l'excellence morale qu'on rencontre dans la vie de tous les jours et dans les témoignages historiques. Pour que l'argument de la rareté extrême par rapport à un trait donné marche, il faut que trois conditions soient remplies : (1) la situation doit être diagnostique (2) l'échantillon des sujets doit être représentatif de la population (3) le pourcentage des sujets qui se comportent de façon moralement condamnable dans la situation doit être extrêmement élevé. A moins qu'on n'ait déjà accepté l'idée de l'imprévisibilité situationnelle radicale du comportement humain et qu'on lui laisse faire implicitement tout le travail explicatif, il y a des difficultés avec les trois points, dont ceux du troisième sont les plus graves.

On vient de voir que le caractère diagnostic des situations est contestable dans la plupart des expériences situationnistes sur la morale, et si la situation n'est pas diagnostique d'un trait, il n'y a aucune preuve de son absence. Dans quelle mesure l'échantillon des sujets est-il représentatif ? L'expérience de Milgram nous donne de bonnes garanties sur ce point, vu qu'elle a été répétée maintes fois avec des gens de tous âges et classes socio-économiques. Mais encore une fois le paradigme de Milgram est le joyau de la littérature situationniste, dans d'autres cas la représentativité est beaucoup plus contestable⁴⁴⁷. Quoi qu'il en soit, le problème central de cette

⁴⁴⁶ Cela ne signifie pas que le théoricien de la vertu puisse s'en sortir si facilement : certaines réinterprétations sont facilement testables et falsifiables. Par exemple, les « obéissants » dans le paradigme de Milgram n'étaient pas forcément très obéissants ou soumis à l'autorité en général dans la vie. Toutefois, l'expérience de Milgram est la plus favorable à l'argumentaire situationniste : c'est plus difficile de « sauver » d'autres expériences situationnistes d'une réinterprétation de ce genre.

⁴⁴⁷ Par exemple, pour des raisons éthiques évidentes le paradigme de la prison de Zimbardo, dans lequel d'ailleurs la tranche d'âge entre 17 et 30 ans était la seule à être représentée, n'a pas été répliqué. Il existe toutefois une expérience de préparation conduite toujours par Zimbardo au printemps de 1971 avec un groupe mixte d'étudiants

interprétation est que le pourcentage de gens qui se comportent de façon éthiquement correcte est souvent suffisamment élevé pour nier que le trait concerné soit « très rare » dans la population. Certes, on risque d'entrer dans des évaluations quantitatives difficiles à trancher d'une façon non arbitraire. Pourtant, il me semble qu'un tiers de sujets non-obéissants dans le paradigme de Milgram soit beaucoup trop pour parler de rareté extrême. Quand le résultat est le plus favorable au situationniste, souvent l'effet n'a pas été répliqué ou le caractère diagnostique de la situation est particulièrement douteux. Dans l'expérience originelle d'Isen et Levin seul un participant sur 25 avait aidé à ramasser les papiers dans la rue après ne pas avoir trouvé de la monnaie dans une cabine téléphonique, mais, comme nous l'avons remarqué, la situation est très peu diagnostique. En plus, le résultat n'a pas été répliqué dans une variante où au lieu des papiers il y avait des paquets⁴⁴⁸

et étudiantes. Les resultants avaient été semblables, P. G. Zimbardo, "Transforming experimental research into advocacy for social change," in *Applying social psychology: Implications for research, practice, and training*, édition M. Deutsch H. A. Hornstein (New York: Wiley, 1975), 37. Il y a aussi deux quasi-répliques, mais avec des normes beaucoup plus strictes qui interdisaient tout acte de violence et d'humiliation, voir S. H. Lovibond, Mithiran et W. G. Adams, "EFFECTS OF 3 EXPERIMENTAL PRISON ENVIRONMENTS ON THE BEHAVIOR OF NONCONVICT VOLUNTEER SUBJECTS," *Australian Psychologist* 14, no. 3 (1979). S. Reicher et S. A. Haslam, "Rethinking the psychology of tyranny" (2004). Dans cette dernière expérience il y a eu une rébellion des prisonniers qui ont pris le contrôle. Par contre, il y a une littérature expérimentale copieuse sur des jeux de rôle dans lesquels les sujets peuvent être amenés à des comportements non-éthiques moins graves. Mais ici on retrouve le problème posé par le caractère diagnostique douteux d'actes qui ne sont certainement pas « gentils », mais peuvent paraître en partie justifiés par le jeu que le sujet est en train de jouer. Sur ce point voir C. L. Doyle, "Interpersonal dynamics in role playing," *American Psychologist* 30 (1975). N. J. Orlando, "The mock ward: A study in simulation.," in *Behavior disorders: Perspectives and trends*, édition O. Milton et R. G. Wahler (Philadelphia: Lippincott, 1973). Vranas défend avec vigueur la représentativité de l'étude de Zimbardo en soulignant que les participants avaient été choisis sur la base d'une absence d'attitudes anti-sociales ou d'autres signes d'une personnalité violente. Il s'agissait donc de personnes normales et pacifiques. Vu le résultat, Vranas pense pouvoir en conclure que tout le monde aurait fait la même chose, voire pire. Même en concédant l'argument, Vranas admet implicitement la rareté d'expériences qui soient aussi probantes que celles de Milgram et Zimbardo : « ...je crois que mon argument a une grande force inductive, parce qu'il suggère qu'avec plus d'intelligence et moins de soucis éthiques (associés à moins de restrictions par des comités institutionnels de contrôle pour la protection des participants humains) les psychologues sociaux pourraient étendre indéfiniment la liste de telles situations ». Vranas, "The indeterminacy paradox: character evaluations and human psychology", 4. Il n'en reste pas moins qu'il y a relativement peu d'expériences *directement* et clairement incompatibles avec la présence d'un trait donné dans la plupart de la population.

⁴⁴⁸ G. A. Blevins et T. Murphy, "FEELING GOOD AND HELPING - FURTHER PHONEBOOTH FINDINGS," *Psychological Reports* 34, no. 1 (1974).

et a été reproduit dans certains essais et pas dans d'autres dans une variante où il s'agissait de poster une lettre affranchie « oubliée » dans la cabine⁴⁴⁹.

Doris remarque qu'il ne s'agit pas de véritables tentatives de répliques, car, dans une approche situationniste, passer de papier à paquet ou à lettre peut tout changer et éliminer l'effet de l'imprévisibilité situationnelle découverte dans l'expérience originelle⁴⁵⁰. Il n'en reste pas moins que l'absence (sinon le manque) de réplique affaiblit l'évidence cumulative de cas où des vertus seraient très rares dans la population.

Des cas plus favorables peuvent être repérés dans la littérature sur l'effet du bon samaritain et sur la dilution de la responsabilité : seuls 10% des séminaristes dans le groupe qui été « très pressé » se sont arrêtés pour aider la personne en détresse et seuls 7% des participants à l'expérience de Latané et Rodin (sur l'urgence signalée par du bruit et des cris dans une pièce à côté) sont intervenus lorsqu'ils se trouvaient en couple avec une personne inconnue (en réalité un collaborateur)⁴⁵¹. Cependant, si l'existence de ces effets a été répliquée de façon satisfaisante, l'extrême rareté du comportement normativement requis n'est pas attestée de manière régulière. Par exemple, dans une expérience les participants étaient seuls dans une cabine mais croyaient être en contact par interphone (pour préserver l'anonymat) soit avec une seule personne, soit avec un groupe de discussion de 5 personnes. Lorsque l'individu à l'autre bout de l'interphone simulait une attaque épileptique (il s'agissait en réalité d'un enregistrement) 100% des sujets qui se croient seuls intervenaient, alors que seulement 62% des participants « en groupe » le faisaient⁴⁵². On voit qu'ici le pourcentage de ceux qui se comportent « bien » est tout à fait respectable. En conclusion, l'idée selon laquelle les expériences situationnistes concernant la morale sont *directement* incompatibles avec une diffusion de traits de caractère plus que marginale n'est pas convaincante. Les preuves du pouvoir de la situation se concentrent de façon très biaisées sur certains aspects du comportement moral (la compassion, l'attitude à aider autrui), se réfèrent souvent à des situations non diagnostiques et ne montrent pas de façon claire la rareté d'attitudes moralement acceptables. Des conclusions situationnistes en mesure de menacer la théorie de la vertu passent toujours par l'acceptation d'une thèse générale sur l'imprévisibilité situationnelle de comportement humain : la discussion des données relatives à la morale n'y ajoute (ni enlève) pas grand chose.

⁴⁴⁹ J. Weyant et R. D. Clark, "DIMES AND HELPING - OTHER SIDE OF COIN," *Personality and Social Psychology Bulletin* 3, no. 1 (1977).

⁴⁵⁰ Doris, *Lack of character: personality and moral behavior* 180.

⁴⁵¹ Latané et Rodin, "A LADY IN DISTRESS - INHIBITING EFFECTS OF FRIENDS AND STRANGERS ON BYSTANDER INTERVENTION."

⁴⁵² Latané et Darley, *The unresponsive bystander: why doesn't he help?*

L'étiologie de l'erreur : l'erreur fondamentale d'attribution

En plus des arguments directs en faveur de la rareté de traits de caractères, les situationnistes ont élaboré une explication des raisons pour lesquelles on se méprendrait à ce point en croyant en l'existence même de structures de personnalité générales et cohérentes. L'étiologie de l'erreur cherche à montrer qu'on n'entretient pas une croyance parce qu'elle est vraie, mais pour d'autres raisons compatibles avec sa fausseté. Si une histoire alternative sur l'origine de nos convictions paraît cohérente et crédible, on a un argument supplémentaire pour les rejeter qui va s'ajouter à toute objection directe à leur validité. Si l'histoire alternative est décisive, il est même possible qu'elle joue le rôle central dans notre motivation à changer d'opinion. Dans le cas du situationnisme, il se trouve qu'on n'est pas confronté au problème d'évaluer le poids comparatif d'une étiologie de l'erreur séduisante face à d'autres considérations, car – nous allons le montrer – les étiologies qui ont été formulées ne sont pas convaincantes. Cette tentative trouve donc sa place dans la liste des interprétations incorrectes du vrai argument situationniste. L'étiologie proposée par les situationnistes n'est pas compatible avec l'analyse de la théorie implicite du caractère du sens commun qu'on a articulée dans un chapitre précédent. Néanmoins, les positions des situationnistes occupent une place centrale dans le débat : nous allons chercher à les articuler et à les réfuter dans le détail.

La stratégie des situationnistes se base sur la littérature sur l'erreur fondamentale d'attribution. Ce dernier est l'un des biais les plus discutés en psychologie sociale, mais il est difficile d'en donner une simple formulation qui soit neutre. Une bonne approximation difficilement contestable de l'erreur mise en évidence dans les expériences est la suivante : les gens ont une tendance à faire des attributions « individuanes » injustifiées à des inconnus. Par « attribution individuelle » on entend l'attribution d'un trait à des niveaux différents des valeurs moyennes dans une population. Autrement dit, on fait une attribution individuelle lorsqu'on croit pouvoir conclure qu'un individu donné est différent d'un individu lambda quelconque selon une certaine dimension évaluative. Dans la pratique, la plupart des attributions sont de ce genre : il est possible que l'on se rende compte qu'un ami est juste aussi courageux que tout le monde, mais normalement lorsqu'on dit « il est courageux » on entend « il est plus courageux que la moyenne ».

Faire des attributions individuelles qui ne sont pas justifiées par les données comportementales dont on dispose est une façon de stéréotyper, de sauter à des conclusions infondées sur les caractéristiques d'une personne : après des décennies d'expériences, il est hors de doute que nous

sommes coupables de ce genre d'erreur⁴⁵³. Mais quel est le lien entre cela et l'étiologie de notre croyance en l'existence de traits de caractère ? La stratégie situationniste sur cette question relie trois niveaux qui ne doivent pas être confondus : d'abord, la littérature situationniste documente un ou plusieurs genres d'erreurs objectives qu'on fait sur une base régulière. Ils sont incontestables d'un point de vue normatif car il s'agit d'erreurs évidentes de prédictions (ex. : les gens croient que presque personne ne va collaborer dans le paradigme de Milgram) ou d'attributions (ex. : on considère que quelqu'un qui n'a fait que recopier et lire à haute voix des arguments pro-Castro est plus pro-Castro que la moyenne)⁴⁵⁴. Cependant, à ce stade les erreurs sont encore non-interprétées, on n'a pas le droit de les associer de façon informative dans un phénomène unitaire sans donner des arguments. Deuxièmement, il s'agit donc de proposer des explications psychologiques plus profondes de ces erreurs. Finalement, et c'est plus important, il faut montrer que les mécanismes psychologiques mobilisés sont suffisants pour comprendre pourquoi on croit à l'existence des traits de caractère. Globalement, il s'agit d'identifier des biais cognitifs suffisants pour expliquer aussi bien des erreurs de prédiction et d'attribution et la croyance en l'existence d'un caractère très cohérent.

Cette stratégie déployée sur trois niveaux doit être articulée indépendamment pour deux phases : il faut rendre compte en termes de biais cognitifs (1) de la *formation* et après (2) du *maintien* d'explications et attributions de traits de caractère par le sens commun. Il est évident que le problème essentiel sera d'expliquer le maintien : aussi bien d'un point de vue intuitif que selon les résultats d'une longue tradition de recherche, il est certain qu'on caricature et stéréotype les gens qu'on ne connaît pas bien et avec lesquels on n'a eu qu'une ou deux rencontres. Par contre, affirmer qu'on ne fait qu'une caricature en termes de traits de personnalité de gens proches qu'on connaît très bien est plutôt fort.

Les explications fournies pour les deux phases mobilisent des biais cognitifs différents. Un ou plusieurs biais seraient à la base de l'attribution injustifiée et des « biais de confirmation » (on évite ou ignore des données contraires à notre impression initiale) expliqueraient le maintien.

⁴⁵³ Ross et Nisbett, *The person and the situation: perspectives of social psychology*.

⁴⁵⁴ D'un point de vue normatif, il ne s'agit peut-être pas d'erreurs. Localement, des erreurs peuvent être une conséquence nécessaire d'une allocation optimale des ressources cognitives. Rechercher à tout prix l'exactitude descriptive peut être contre-productif. Dans cette hypothèse, faire ce type d'erreur dans les contextes appropriés pourrait être un mal nécessaire, étant donné la limitation des ressources disponibles. Cependant, ici nous sommes intéressés à une question descriptive (l'existence et la détection de certaines structures psychologique), non pas d'utilité pragmatique. De ce point de vue, il s'agit d'erreurs incontestables.

Cette séparation nous facilite la tâche car la première question est beaucoup plus complexe : comprendre les mécanismes cognitifs qui expliquent l'attribution biaisée de traits revient à donner un modèle d'une grande partie de la cognition sociale ! Il existe une littérature énorme et beaucoup de controverses à ce sujet en psychologie : heureusement pour les besoins de la présente discussion nous n'avons pas trop à nous soucier de ces détails fort intéressants, car nous sommes déjà assurés du résultat (quelle que soit la cause, il est certain qu'on tend à caricaturer les gens) et le problème plus important du maintien des stéréotypes au fil de l'approfondissement d'une connaissance se révèle indépendant et plus facile à résoudre. Néanmoins, commençons par la première question, qui reste instructive pour comprendre le rôle de l'attribution de traits de caractère dans l'économie de la cognition humaine et a été discutée dans la littérature philosophique sur le situationnisme.

Formation d'impressions stéréotypées

Parmi les philosophes, Harman est le premier à souscrire à une étiologie de l'erreur de la formation d'impressions⁴⁵⁵. Il développe une idée formulée par Ross et Nisbett par rapport à ce qui est saillant et ce qui ne l'est pas dans une scène d'action :

« Quand on fait la distinction entre une figure et l'arrière-plan, on prête plus d'attention à la figure qu'à l'arrière-plan et on cherche à expliquer ce qui se passe en termes de caractéristiques de la figure plutôt que de l'arrière-plan. Typiquement, l'acteur est la figure et la situation est l'arrière-

⁴⁵⁵ La classification situationniste de l'ensemble des erreurs qui doivent trouver une explication (premier niveau) est contestable. Les expériences mettent en évidence des erreurs de prédiction (les gens n'anticipent pas le résultat de Milgram) et d'attribution injustifiée. Les situationnistes réduisent toutes ces erreurs à des cas du deuxième genre. Ils considèrent que l'on n'anticipe pas l'obéissance dans le paradigme de Milgram pour la raison suivante : (1) comme on est hâtif dans l'attribution des traits, si on voyait une personne « obéir » on penserait qu'elle est très cruelle (2) étant donné que l'on ne croit pas que la plupart des gens soient cruels, on en conclut que la plupart des gens ne vont pas collaborer. Le même schéma serait valide pour les autres paradigmes expérimentaux : si quelqu'un ne s'arrête pas pour assister une personne en détresse, c'est qu'il est insensible (attribution hâtive) ; on ne croit pas que la plupart des gens soient insensibles, donc la plupart des gens vont s'arrêter. Le problème de cette explication n'est pas qu'elle manque de plausibilité, mais plutôt qu'elle est donnée pour acquise par certains situationnistes, au niveau même de la classification des erreurs cognitives qu'il faudrait expliquer. Voir L. D. Ross, "Getting down to fundamentals: Lay dispositionism and the attributions of psychologists," *Psychological Inquiry* 12, no. 1 (2001). Or, sans preuves ultérieures, il s'agit là de pure spéculation. Il n'est pas évident que l'erreur de prédiction soit entièrement expliquée par l'erreur d'attribution.

plan, ainsi on cherche à expliquer l'action par les propriétés de l'acteur sur la scène plutôt que celles de la situation en arrière-plan⁴⁵⁶. »

Dans un article récent, Webber remarque que les données censées étayer le fait que l'attribution de traits suit le facteur le plus saillant dans le contexte ne sont pas pertinentes⁴⁵⁷. Il s'agit d'une série d'expériences citées par Ross et Nisbett où il paraît que l'attribution de responsabilité causale à l'acteur ou à l'environnement varie en fonction de la capacité qu'ont les deux d'attirer l'attention⁴⁵⁸. Suivant que l'environnement est statique ou en mouvement et l'acteur immobile ou en train de bouger (ou encore habillé de façon voyante) on va juger que ces deux facteurs sont plus ou moins influents. Seulement, à y regarder de près, les résultats des expériences citées ne vont pas tous dans la même direction et – plus important – ils sont en tout cas non pertinents.

Dans les expériences, on avait demandé aux sujets d'évaluer l'importance de la situation et de l'agent sur deux échelles indépendantes (l'augmentation d'une valeur n'entraîne pas la diminution de l'autre). Or, la seule signification cohérente de la différence « explication situationnelle/explication dispositionnelle » dans le contexte de l'attribution d'un trait de caractère est celle qui oppose des actions individuantes (= explication dispositionnelle) qui justifient l'attribution de propriétés spécifiques à l'individu et des actions non individuantes (= explication situationnelle) qui ne justifient pas de conclure que l'agent a un profil différent par rapport à la moyenne de la population. « C'est la situation » veut dire « pour autant que je sache, c'est un gars comme tous les autres », « c'est la personne » veut dire « je crois qu'il est (disons) courageux ».

Ces deux genres de styles explicatifs s'opposent clairement sur la même échelle, si c'est la situation cela veut dire que ce n'est pas la personne. Il ne peut pas s'agir du même sens de la distinction situation/personne qui est mobilisée dans les expériences, où les échelles sont indépendantes. Après notre discussion ci-dessus, il est clair qu'une interprétation causale de l'opposition n'a pas de sens :

⁴⁵⁶ Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," 325.

⁴⁵⁷ Jonathan Webber, "Character, Common-Sense, and Expertise," *Ethical Theory and Moral Practice* publié online (2006).

⁴⁵⁸ Ross et Nisbett, *The person and the situation: perspectives of social psychology* 140. R. M. Arkin et S. Duval, "FOCUS OF ATTENTION AND CAUSAL ATTRIBUTIONS OF ACTORS AND OBSERVERS," *Journal of Experimental Social Psychology* 11, no. 5 (1975). L. Z. McArthur et D. L. Post, "FIGURAL EMPHASIS AND PERSON PERCEPTION," *Journal of Experimental Social Psychology* 13, no. 6 (1977). S. E. Taylor et S. T. Fiske, "POINT OF VIEW AND PERCEPTIONS OF CAUSALITY," *Journal of Personality and Social Psychology* 32, no. 3 (1975).

les stimuli de la situation et les dispositions de l'agent sont à 100% nécessaires et importants pour expliquer toute action⁴⁵⁹.

S'il n'est pas plausible que les biais d'attribution et la formation d'impressions stéréotypées soient expliqués par le facteur qui paraît le plus saillant entre l'agent et le contexte, il y a toute une littérature qui élabore des solutions plus prometteuses que ce ne serait pas pertinent d'explorer ici⁴⁶⁰.

Le maintien d'impressions stéréotypées dans l'approfondissement d'une connaissance

Quel que soit le mécanisme qui explique notre tendance à stéréotyper et attribuer des traits de façon hâtive, une chose est certaine et prouvée au-delà de tout doute raisonnable : nous sommes massivement coupables de ce genre d'erreurs d'attribution. Vu qu'il s'agit d'erreurs, une explication satisfaisante du pourquoi on les fait *doit* forcément exister et le situationniste ne se trouve pas dans une position difficile s'il ne peut indiquer laquelle avec certitude. La vraie question sera d'expliquer le maintien de stéréotypes face à des données contradictoires au fil de la connaissance d'une personne. La stratégie situationniste consiste à montrer que contrairement à ce qu'on pourrait croire, nous voyons nos proches dans un nombre bien limité de situations, et donc les pièges situationnels qui pourraient manifester l'incohérence de leur conduite sont tenus à l'écart par la structure située et définie de la relation : il y a là un biais de *sélection des situations*

(b) on a un *biais de confirmation* qui consiste à écarter sélectivement toute donnée susceptible de mettre à mal l'image stéréotypée qu'on s'est faite d'une personne dès les premières rencontres.

En ce qui concerne le biais de sélection des situations, les situationnistes avancent trois genres d'arguments⁴⁶¹ :

⁴⁵⁹ On peut se demander ce que les participants ont à l'esprit lorsqu'ils répondent à des consignes formulées en des termes si confus. Les expressions « causalité », « responsabilité causale » etc. sont célèbres pour être polysémiques et se réfèrent vraisemblablement à une pluralité de concepts différents dans l'usage ordinaire. Le fait qu'il y ait des résultats répliquables ne signifie pas que les conclusions soient significatives et représentatives de la position du sens commun sur la notion de causalité qui nous intéresse, car il peut y avoir une tendance quasi universelle à réagir de façon bien déterminée face à des questions ambiguës.

⁴⁶⁰ Voir Gilbert et Malone, "THE CORRESPONDENCE BIAS."

⁴⁶¹ Doris, *Lack of character: personality and moral behavior* 102-104.

1. à bien y réfléchir les relations, même très longues et très intimes, se structurent sur un ensemble plutôt limité de contextes : il y a beaucoup de situations dans lesquelles on ne voit jamais un ami ou un parent, à commencer bien sûr par toutes les situations où on n'est pas là (on ne voit jamais un partenaire avec ses amis ou sa famille mais sans nous)
2. les dynamiques de la vie sont telles qu'on n'est pas confronté à des faits réels ou contrefactuels qui iraient infirmer nos jugements sociaux précédents : dans tout « choix » d'une personne, que ce soit en amour ou dans le cadre d'une sélection d'embauche, on ne peut par définition savoir comment se seraient débrouillés les gens écartés, et face à un échec, on ne peut jamais exclure que d'autres options nous auraient également amené à la faillite.
3. il y aurait des effets d'attentes auto-réalisatrices : on ne se conduit pas de la même façon avec quelqu'un qu'on nous annonce être timide (ou extrovertis). Vraisemblablement, on va anticiper et faciliter les réactions auxquelles on s'attend. De cette manière, même les stéréotypes les plus grossiers sur les caractères nationaux peuvent s'auto-réaliser : si dans son imaginaire on rêve d'un pays, on va avoir une attitude spéciale qui va stimuler une réaction du même genre chez les habitants de ce pays, qui dans l'absolu n'ont rien de plus que ses propres compatriotes.

Le deuxième argument n'est pas pertinent, car nous avons déjà concédé le caractère stéréotypé des jugements qu'on porte sur des inconnus, et nos proches ne sont certainement pas des personnes qu'on ne recroise plus dans notre vie.

Le premier point sous-évalue la quantité d'information indirecte qu'on arrive à accumuler par le témoignage d'autrui : s'il est vrai que des contextes restent bien isolés aussi et surtout d'un point de vue informationnel (ex. : des parents qui ne savent pas ce que font les fils adolescents) il faut se souvenir que Doris et Harman postulent un potentiel colossal d'incohérence diffusée de façon homogène dans l'espace des situations de la vie courante : il ne s'agit pas de cacher à nos parents des habitudes circonscrites. Il est difficile d'imaginer que la structure bien définie d'une relation et les limitations par rapport à ce qu'on peut découvrir par d'autres biais soient suffisantes pour tenir à l'écart cette masse volumineuse de données incohérentes.

Il y a une bonne quantité d'études qui attestent l'existence des phénomènes mis en évidence par le troisième argument⁴⁶². Les expériences dans lesquelles une manipulation des attentes des professeurs (au hasard, dans un bon ou mauvais sens) préalable à la rencontre avec un nouveau

⁴⁶² Pour une synthèse, voir Mark Snyder et Olivier Klein, "Construing and constructing others: On the reality and the generality of the behavioral confirmation scenario," *Interaction Studies* 6 (2005).

étudiant ont conditionné dans le même sens les résultats scolaires effectifs sont célèbres : il s'agit de l'effet Pygmalion⁴⁶³. A part cela, l'autre paradigme (de laboratoire) typique qui a été développé au fil de plusieurs décennies prévoit trois participants : le « percevant », la cible et le juge indépendant. Les attentes du percevant sur la cible sont manipulées, on va lui dire que la cible est extrovertie, sympathique ou antipathique (etc.) selon les cas et au hasard, sans aucun lien avec les vraies attitudes qui peuvent donc correspondre ou non. Les deux, qui ne se connaissent pas et ne prévoient pas de se revoir après, ont une brève discussion au terme de laquelle le niveau du trait concerné chez la cible est jugé aussi bien par le percevant que par un juge indépendant qui ne connaît pas les attentes du percevant et ne fait que visionner l'enregistrement vidéo. On parle de « confirmation comportementale » (*behavioral confirmation*) lorsque le verdict du juge va dans le même sens que les attentes, et de confirmation perceptive (*perceptual confirmation*) lorsque le jugement du percevant confirme ses attentes. Autrement, il s'agirait respectivement de réfutation comportementale et réfutation perceptive.

Dans ce paradigme, des effets de confirmation comportementale et perceptive sont produits de façon fiable. Il y a toutefois une limitation importante : la plupart des expériences sont, comme nous venons de le dire, des échanges d'inconnus qui ne prévoient pas de se revoir, et cela n'est pas une situation représentative du phénomène qui nous intéresse, c'est-à-dire la capacité à produire de la cohérence dans l'attitude d'un partenaire par des attentes auto-réalisatrices dans le cadre d'une relation continue, vu que par définition on prévoit de revoir ses proches.

Cependant, il y a des expériences où on a dépassé le cadre de la rencontre entre inconnus : le résultat d'une anticipation de la possibilité de rencontres ultérieures est de renforcer l'effet. Si vous pensez qu'il est possible de revoir une personne, votre capacité d'influencer son attitude sera plus importante⁴⁶⁴. D'autres expériences démontrent qu'au fil de quelques rencontres avec le même percevant la cible peut insister dans l'attitude induite et la confiance du percevant dans son jugement va augmenter (même si l'attitude induite initialement n'était pas censée correspondre à la cible).

Cela semble aller dans la bonne direction pour le situationniste, car ce serait une preuve de la capacité à induire et stabiliser une attitude absente chez quelqu'un, en se convaincant de plus en plus que le trait concerné le caractérisait depuis le début : c'est justement le genre de mécanisme qui

⁴⁶³ Pour une synthèse de la littérature sur cet effet, voir R. Rosenthal et D. B. Rubin, "Interpersonal expectancy effects: the first 345 studies", *Behavioral and Brain Science* 3 (1978).

⁴⁶⁴ Cela ne semble marcher que pour les attentes positives et non pour les attentes d'attitudes indésirables, voir A. A. Stukas et M. Snyder, "The influence of possible future interactions on first encounters" (non publié: La Trobe University, 2003).

pourrait expliquer la difficulté à rencontrer des situations contradictoires. On ferait en sorte qu'un proche soit *vraiment* courageux ou honnête, mais seulement avec *nous* et grâce à notre capacité à induire le trait. Cependant, il faut remarquer qu'en l'absence d'études qui suivent ce processus d'induction après la toute première série de rencontres, penser que ses effets ne sont pas contrecarrés ou balancés par l'ensemble des autres processus qui interviennent dans l'approfondissement d'une connaissance au fil des années relève de la pure spéculation. Encore une fois, il est tout à fait plausible que l'effet marche dans le cas de quelques traits : notre présence peut rendre courageux et sensible un ami en écartant de notre regard le potentiel d'incohérence qu'il pourrait exprimer dans ces aspects de sa vie, mais il est plus difficile de croire que notre contact va imposer une entière panoplie de traits pour empêcher que nous venions en contact avec son incohérence fondamentale.

Dans l'utilisation situationniste de ces données il y a aussi un problème plus grave que la faiblesse de l'effet. A y regarder de près, il ne s'agit pas d'un phénomène qui avance la cause de l'incohérence du comportement humain, car les données suggèrent la possibilité d'induire vraiment des traits généraux, non seulement des traits locaux qui se manifesteraient en présence de la personne à l'origine de l'influence. Il a été prouvé qu'après une première rencontre la cible a tendance à « intérioriser » l'attitude induite qui va être maintenue dans une deuxième expérience où le nouveau percevant a des attentes différentes⁴⁶⁵. La corrélation entre les deux essais a même atteint 0,9 dans certains cas, ce qui prouve la puissance de l'effet de généralisation⁴⁶⁶. Le risque d'une vraie intériorisation à partir des attentes d'autrui n'est pas seulement un effet marginal documenté dans une littérature psychologique peu connue. Il s'agit au contraire de la raison principale pour laquelle ces phénomènes sont étudiés, car il semble qu'un processus d'intériorisation de stéréotypes négatifs joue un rôle crucial dans la psychologie de la discrimination.

Il ne sera pas possible de couvrir cette littérature ici, mais les exemples se multiplient, on va d'expériences où des femmes censées être moins douées en mathématique ont eu des résultats inférieurs à des femmes auxquelles on avait expliqué l'égalité cognitive entre les sexes⁴⁶⁷, à des

⁴⁶⁵ A. A. Stukas et M. Snyder, "Targets' awareness of expectations and behavioral confirmation in ongoing interactions," *Journal of Experimental Social Psychology* 38 (2002).

⁴⁶⁶ M. Snyder et W. B. Swann, "Behavioral confirmation in social interaction: From social perception to social reality," *Journal of Experimental Social Psychology* 14 (1978).

⁴⁶⁷ Ilan Dar-Nimrod et Steven J. Heine, "Exposure to Scientific Theories Affects Women's Math Performance," *Science* 314, no. 5798 (2006).

analyses qui montrent l'effet pervers de la possibilité de mobilité sociale et de sortie d'une classe défavorisée : dans ce dernier cas, la possibilité de promotion personnelle pour l'individu rend cognitivement très saillant le clivage social à dépasser, stimulant ainsi le maintien des stéréotypes⁴⁶⁸.

Même si un trait de caractère induit de cette façon en un sens ne fait pas partie de la « vraie » personnalité potentielle des victimes, ce n'est pas une question qui doit intéresser le situationniste : son approche prévoit une vulnérabilité généralisée à des pièges situationnistes, et toute forme de cohérence stable dans les attitudes des gens, produite par n'importe quel biais, joue contre lui.

Pour résumer, il ne semble pas crédible qu'il existe des effets de sélection des situations suffisamment forts pour tenir à l'écart l'immense potentiel d'incohérence dans l'attitude de nos proches qui est postulé par les situationnistes. La partie de leur vie qui se passe loin de notre regard – protégée contre des fuites d'information qui pourraient nous rejoindre – est trop localisée et spécifique pour jouer ce rôle de barrage. Notre capacité à influencer leur conduite dans des situations où nous sommes présents – certainement active et importante pour expliquer des jugements stéréotypés et des mythes *locaux* dans l'image que nous nous faisons d'eux (la sympathie proverbiale de notre père, la générosité sans limites de notre meilleur ami) – ne paraît pas suffisamment forte pour alimenter un grand stéréotype global détaché de la réalité. Pire, il semble que, à la limite, on pourrait vraiment arriver à façonner leur caractère avec des effets généraux et stables, si bien que l'incohérence postulée serait effacée.

Le situationniste veut nous convaincre de la puissance des biais de confirmation : on rencontre bien au moins une grande partie des signes de l'incohérence de nos proches, mais on l'ignore en l'écartant de notre attention et ne la retenant pas en mémoire. Là aussi, il y a une littérature volumineuse, mais, comme les exemples cités par les situationnistes le démontrent, plutôt indirecte⁴⁶⁹.

Comme il arrive dans l'évaluation des théories scientifiques, lorsqu'une observation met à mal une théorie enracinée il est tout à fait légitime de douter de l'observation avant d'abandonner une croyance. Ainsi, face à un contre-exemple dans la conduite d'un ami normalement courageux, on a le droit de se demander si le contre-exemple est vraiment tel ou s'il y a une explication différente. Dans d'autres contextes, on semble abuser de cette stratégie de conservatisme épistémique. Il y a plusieurs exemples de maintien d'une opinion ou d'une impression face à des informations censées

⁴⁶⁸ H. Tajfel et J. C. Turner, "The social identity theory of intergroup behavior.," in *The psychology of intergroup relations*, édition W. G. Austin (Chicago: Nelson-Hall, 1986).

⁴⁶⁹ Doris, *Lack of character: personality and moral behavior* 96, 102.

la disqualifier complètement. Dans une expérience très citée, Ross demandait à des sujets de faire le tri entre des vraies et de fausses lettres laissées par des suicidaires⁴⁷⁰. Certains sujets se montraient plus capables que d'autres, mais tout était en effet truqué : les lettres étaient toutes fausses et les notes assignées au hasard. Après une séance de débriefing dans laquelle la structure de l'expérience était expliquée, les sujets qui avaient eu auparavant les meilleurs notes continuaient de s'auto-évaluer (et d'être évalués par des observateurs) comme étant meilleurs à la tâche. Un autre phénomène assez médiatisé est la tendance, après un débat public dans lequel on est confronté à des données et arguments clairs contre son propre point de vue, à renforcer la position initiale (au lieu d'alimenter le doute). Dans le même paradigme, des données contradictoires (deux comptes-rendus avec des conclusions opposées sur la force de dissuasion de la peine capitale) ont le même effet de polarisation⁴⁷¹.

Est-ce que cela se reproduit systématiquement dans le cas des exemples d'incohérence qu'on rencontre et qui devraient mettre à mal nos stéréotypes? Si oui, la puissance du phénomène est-elle suffisante pour défendre l'image stéréotypée de nos proches contre l'assaut d'une quantité colossale d'incohérence? Il n'y a pas de données expérimentales directes, mais une telle ambition explicative sur des bases expérimentales aussi indirectes relèverait de la pure spéculation. Le premier type d'expérience est expliqué vraisemblablement par un mécanisme d'ancrage (*anchoring*) et de correction insuffisante, dans lequel une première impression est formée et continue à exercer son influence tant qu'elle n'est pas complètement effacée par un processus ultérieur, qui demande de l'attention et des énergies mentales supplémentaires. Comme il est facile de l'imaginer, souvent ces ressources viennent à manquer et l'influence continue de s'exercer par défaut de façon inopportune. Or, ce genre d'explication est particulièrement mauvais dans le cadre de la compréhension de nos proches, vu que dans une fréquentation qui se déroule sur plusieurs années on a bien le temps de corriger les premières impressions si elles sont mauvaises. En ce qui concerne la volonté éventuelle de défendre au-delà du raisonnable et contre n'importe quel argument les stéréotypes initiaux – tout comme on défend son propre avis sur la peine capitale –, son existence n'est tout simplement pas prouvée.

⁴⁷⁰ L. Ross, M. R. Lepper et M. Hubbard, "PERSEVERANCE IN SELF-PERCEPTION AND SOCIAL PERCEPTION - BIASED ATTRIBUTIONAL PROCESSES IN DEBRIEFING PARADIGM," *Journal of Personality and Social Psychology* 32, no. 5 (1975).

⁴⁷¹ C. G. Lord, L. Ross et M. R. Lepper, "BIASED ASSIMILATION AND ATTITUDE POLARIZATION - EFFECTS OF PRIOR THEORIES ON SUBSEQUENTLY CONSIDERED EVIDENCE," *Journal of Personality and Social Psychology* 37, no. 11 (1979).

Pour résumer, comme cela a été le cas pour les biais de sélection, la force des biais de confirmation qui pourraient être actifs dans l'interprétation du comportement de nos proches paraît franchement insuffisante pour expliquer la mise à l'écart de la masse de données contradictoires auxquelles on va être confronté si l'hypothèse du situationniste est correcte. Globalement, l'explication situationniste des raisons pour lesquelles on croirait à l'existence des traits de caractère larges (alors que le comportement est massivement fragmenté) ne marche pas : elle n'arrive pas à expliquer pourquoi on croirait que *nos proches* possèdent des traits larges.

Il y a aussi un problème plus profond que tous ceux qu'on vient de traiter : la tentative d'explication de l'ignorance de la fragmentation du caractère ne peut marcher car le phénomène qu'on essaie d'expliquer n'est pas réel. Nous sommes bien conscients des incohérences dans la conduite de nos proches ! Certes, le sens commun prend parfois au sérieux la notion de trait de caractère large, en la faisant figurer dans de véritables explications générales du comportement telles que « la générosité amène à offrir ses propres biens », « le courage pousse à aider les victimes » (etc.), mais si on réfléchit à l'expérience quotidienne avec nos proches, il est évident que nous mobilisons une connaissance plus détaillée, composée par des traits de caractère plus spécifiques. Par exemple, nous savons très bien *a priori* que notre meilleur ami n'est pas timide, mais qu'il le devient avec les jolies femmes ou s'il doit appeler un inconnu au téléphone, que notre père est autoritaire mais qu'il n'oserait pas nous réprimander en présence de son vieil ami d'enfance, que notre grand-père est généreux mais qu'il ne faut pas lui demander d'argent le jour où il paye les impôts, que le professeur est normalement impartial mais qu'il vaut mieux ne pas faire devant lui des commentaires ironiques. Personne ne serait surpris d'entendre un ami faire ce genre de considérations sur ses proches ou les gens qu'il connaît bien, ni de lire une telle description dans un roman. Si la théorie de la personnalité implicite du sens commun se réduisait aux traits larges critiqués par Doris, vu que les gens sont familiers avec l'incohérence dans le comportement de leurs proches, il s'ensuivrait qu'ils devraient penser que tout le monde possède un caractère... à l'exception de leur propre entourage ! Il est évident que la « personnologie » populaire est très complexe : les concepts de traits de caractère sont utilisés de façon différente suivant qu'on parle d'un inconnu ou d'un proche. Dans le premier cas, la notion de (disons) courage est incompatible avec des manifestations d'incohérence – si on était convaincu qu'un étranger n'aiderait pas un enfant, on refuserait de l'appeler courageux – ; dans le deuxième elle est plus souple et semble compatible avec notre connaissance manifeste de l'incohérence de nos proches.

Pour conclure, c'est toute la stratégie d'explication de l'erreur proposée par les situationnistes qui est à rejeter : elle ne marche pas, et ce n'est pas un hasard, car les gens ne sont pas coupables du

genre d'erreur qu'on leur attribue. Implicitement, ils sont bien conscients de la complexité et des irrégularités du comportement humain.

Conclusion

Nous avons analysé plusieurs façons d'interpréter la littérature sur l'influence de la situation. Il en ressort que la menace posée pour la théorie de la vertu se base sur une généralisation extrême d'effets d'influence situationnelle. L'idée que des pièges situationnels nous encerclent systématiquement est incompatible avec la possibilité d'une cohérence large, et n'est pas étayée par une théorie psychologique intéressante et profonde : elle se base uniquement sur le constat qu'on n'a pas pour l'instant trouvé de limites à la possibilité de reproduire et multiplier de tels effets.

La généralisation est aussi problématique à cause du manque d'une bonne explication des raisons pour lesquelles on ne se rendrait pas compte de la non-existence des traits larges, car il semble que nous connaissons bien les incohérences de nos proches. Dans les chapitres suivants, il faudra s'intéresser directement aux mécanismes psychologiques qui pourraient expliquer la présence ou le manque de cohérence, pour pouvoir estimer la force du problème posé par les pièges situationnels et élaborer une stratégie de reconstruction d'une théorie de la vertu empiriquement adéquate.

La vertu et la psychologie de la compétence et de l'expertise

L'étude de la psychologie de la personnalité et de la psychologie sociale nous donne des raisons de croire que la vertu est rare. Cette rareté est beaucoup plus problématique que les théoriciens de la vertu n'ont l'air de le croire : on commence à soupçonner que la vertu ne pourra pas rendre compte de l'excellence morale (dont on a des témoignages directs) et que le développement de la vertu demande des efforts déraisonnables par rapport à la psychologie humaine. Il est temps de s'occuper plus directement du défi psychologique représenté par le développement de la vertu.

L'optimisme quant à la possibilité de devenir vertueux se base sur l'une des thèses centrales de la psychologie morale classique : la vertu est une forme spécifique de compétence (*skill*). Comme on peut maîtriser la technique d'un instrument musical, la conduite d'un navire, l'équitation ou le jeu des échecs, on pourrait, au fil de l'éducation, maîtriser la complexité d'une conduite moralement irrépréhensible. Cependant, une simple identification entre les notions de vertu et de compétence pratique n'est pas dans l'intérêt stratégique de la théorie de la vertu. La vertu n'est pas une compétence quelconque : il s'agit d'une compétence complexe, de très haut niveau et définissant de façon profonde l'identité du sujet qui la possède. On fait de l'équitation mais on *est* courageux : la vertu nous définit.

Malgré le fait que la vertu soit une forme de compétence *sui generis*, la stratégie de la tradition dépend de manière cruciale de ce que les vertus et les compétences ont en commun. Le point de départ des anciens était un débat très large sur la psychologie de la compétence pratique (notamment sur le rôle que le raisonnement, les émotions, la perception et la pratique y jouent), l'interprétation de la notion de vertu dépendant entièrement de ce corpus de réflexion. La compétence est la notion connue, l'*explicans* : la vertu la notion à expliquer, l'*explicandum*.

D'un point de vue technique, la psychologie morale de la vertu est une théorie des traits conçus comme des compétences de haut niveau. Des deux aspects fondamentaux de la théorie, (1) la dynamique et l'interaction entre les vertus (comment le courage a besoin aussi de la justice, l'honnêteté de la patience, etc.) et (2) la structure psychologique de chaque trait (dans quelle mesure le courage est une question de perception, de délibération rationnelle, de réactions émotionnelles correctes etc.), le deuxième conditionne largement le premier. Le type d'interaction nécessaire ou souhaitable dépend clairement des propriétés psychologiques des processus censés interagir. Si bien que les fondamentaux d'une psychologie de l'expertise constituent la base de l'édifice théorique. Notamment, nous verrons que l'optimisme classique sur l'apprentissage de la vertu se justifie par

l'idée selon laquelle les vertus se développent et se complètent entre elles de la même manière que les compétences : autrement dit, l'optimisme des anciens n'était pas basé sur ce que les vertus auraient de spécial, mais sur ce qu'elles auraient en commun avec toute forme de compétence pratique.

Voici la ruse stratégique : faire passer le côté plus sophistiqué, haut niveau et philosophiquement intéressant de la vertu dans la même enveloppe que le côté standard d'une compétence commune. Mais le tour de passe-passe ne marche pas. Dans ce chapitre, nous montrerons la distinction que la psychologie contemporaine opère entre une simple compétence et une véritable expertise dans une tâche. Si la psychologie de la compétence répond aux attentes de la théorie de la vertu, la psychologie de l'excellence (ou expertise) a des caractéristiques incompatibles avec l'interaction entre vertus qui est envisagée par la tradition. Or, malheureusement, la vertu étant très rare, elle relève sans doute de la psychologie de l'excellence, non pas de la simple compétence.

Il est possible d'atteindre un niveau moyen et acceptable de compétence avec une quantité raisonnable d'effort. En outre, les compétences sont en général cumulatives et coopératives : si je veux m'inscrire à une université espagnole en mathématique, être bon en espagnol et aussi en mathématique vont m'aider conjointement. Par contre, l'excellence véritable (être parmi les meilleurs) demande un investissement énorme en termes de temps et d'entraînement pointu (pas simplement de pratique), et produit des véritables adaptations psychophysiques profondes à la tâche. L'expert est un véritable « mutant fonctionnel ».

Des niveaux d'excellence dans plusieurs domaines ne s'accumulent et ne se soutiennent pas réciproquement pour deux raisons : (1) des limites logistiques et temporelles, car il faut énormément de temps et d'entraînement pointu dans la tâche spécifique pour atteindre un niveau d'excellence (contrairement à la compétence, il n'y a pas de transfert ou d'apprentissage accéléré entre des domaines d'excellence proches) ; (2) l'adaptation psychophysique profonde à la tâche prédispose à de bons niveaux de *compétence* dans d'autres tâches proches, mais enlève du potentiel et de la flexibilité pour les autres adaptations profondes qui seraient nécessaire pour atteindre l'expertise. L'excellence est atteinte très souvent dans un seul domaine, et exceptionnellement dans deux ou trois domaines liés et complémentaires, mais pas plus. À partir du constat qu'un individu est excellent dans une ou deux tâches, on peut s'attendre à découvrir des points faibles dans d'autres domaines, plutôt qu'un profil sans faille.

Une autre stratégie qui soutient l'optimisme des théoriciens de la vertu – et qui est mise à mal par la distinction entre compétence et excellence – est la tendance à minimiser la partie de délibération et réflexion explicite et à maximiser la partie de perception et automatisme dans le fonctionnement psychologique de l'individu vertueux. Tout ce qui peut être intériorisé de façon profonde et passer

au « pilote automatique » libère des ressources cognitives précieuses pour d'autres tâches : cela paraît évident dans l'expérience ordinaire et a été confirmé par la psychologie contemporaine⁴⁷². Ainsi, plus on peut automatiser, plus on peut exécuter des tâches complexes et difficiles. D'où l'intérêt à ce qu'un maximum de la psychologie de la vertu relève de l'automatisme, étant donné les compétences extrêmement ambitieuses que la notion de vertu postule. Nous verrons que le maximum d'automatisation est atteint au niveau de la compétence, alors qu'au niveau de l'expertise, l'analyse et la réflexion explicite retrouvent leur place.

Psychologie de la compétence et de l'expertise

Un ensemble aussi vaste et hétéroclite de tâches comme jouer aux échecs, pratiquer une activité sportive, faire des mots croisés, faire des diagnostics corrects à partir d'une radiographie, conduire une voiture, élaborer des plans d'intervention en politique agricole (bref, pratiquement n'importe quelle tâche) sont caractérisées par des régularités impressionnantes au niveau (1) du genre d'apprentissage nécessaire pour progresser à de différentes phases ; (2) du temps nécessaire pour la progression ; (3) de la psychologie associée aux différents stades de maîtrise. Avant d'analyser les données, une telle généralité dans la littérature paraît difficilement crédible à quelqu'un armé d'un sain scepticisme scientifique. Pourtant, les conclusions de plus d'un siècle de recherche sont claires. Dans l'exercice de n'importe quelle activité, pratique ou intellectuelle qu'elle soit, il y a deux objectifs saillants qui marquent des tournants dans la courbe d'apprentissage : la compétence (ou maîtrise), et l'excellence. La compétence correspond intuitivement à un niveau acceptable et moyen de maîtrise d'une tâche : en d'autres termes, savoir vraiment faire quelque chose. Être un bon joueur d'échec, être diplômé en violon, savoir vraiment jouer au poker, avoir la capacité de diagnostic d'un médecin pratiquant moyen, etc. L'excellence est définie par l'appartenance au groupe des meilleurs pratiquants d'une activité : non nécessairement les dix champions les plus connus, mais il s'agit de très haut niveau.

La durée du processus de développement de la maîtrise est très variable par rapport à la difficulté de la tâche : il peut s'agir de quelques heures ou de quelques mois pour des simples tâches motrices, jusqu'à des années pour développer la capacité de diagnostic d'un médecin moyen. Dans les deux

⁴⁷² P. L. Ackerman, "New Developments in Understanding Skilled Performance," *Current Directions in Psychological Science* 16, no. 5 (2007).

cas, l'acquisition de la maîtrise est essentiellement un processus d'automatisation : dans la courbe d'apprentissage, la maîtrise représente le pic d' « automatisation relative »⁴⁷³.

Bien sûr, la maîtrise ne correspond pas à un niveau fixé de façon très précise, il s'agit d'un intervalle très large. Contrairement à l'opinion reçue selon laquelle ceux qui possèdent un niveau moyen de compétence se vaudraient à peu près tous (le diplômé moyen de violon, l'automobiliste moyen) et il y aurait, par contre, plusieurs styles de champions, en réalité on retrouve beaucoup plus de variabilité au niveau de la maîtrise que de l'excellence⁴⁷⁴. En ce qui concerne la première, la pratique a un effet de nivellement des performances uniquement pour des tâches faciles ou à la structure bien délimitée. Lorsque cela n'est pas le cas, un écart se creuse entre les meilleurs et les moins bons. Cet écart peut être prédit par une série limitée de facteurs, suivant la nature essentiellement motrice ou intellectuelle de la tâche.

Une tâche est facile si le novice atteint un niveau raisonnable de succès depuis le début, serait-ce lentement et au prix de pas mal d'effort. Par exemple, on peut arriver rapidement à taper sur un clavier avec les deux mains, alors que jouer le violon décemment est impossible au tout début de l'entraînement. Dans des tâches motrices comme taper sur un clavier ou dans des tâches intellectuelles bien délimitées (telles qu'apprendre la méthode pour faire des divisions), la quantité de pratique est le meilleur indice de la performance. Les différences de talent naturel s'estompent : tout se passe comme si – l'objectif et la procédure étant bien clairs – tout le monde au bout d'un certain temps trouvait la façon de les implémenter à partir de ses propres ressources.

Si la tâche est plus complexe et ouverte – par exemple, jouer à un jeu vidéo qui demande beaucoup de capacités ou être bon en mathématiques en général – l'écart se creuse rapidement entre ceux qui bénéficient de la pratique et ceux qui restent derrière. Cet effet se produit en particulier quand, pour avoir un succès, il faudrait maîtriser en même temps plusieurs compétences hétérogènes, depuis le début. Autrement dit, lorsque la tâche est composée par un ensemble de sous-tâches. Un jeu vidéo peut demander une bonne coordination motrice, du sens de l'orientation, une bonne mémoire, etc. Lorsque la sous-division en tâches plus simples est nette, un apprentissage séparé de chaque tâche peut faire progresser le novice qui n'est pas doué. La progression est plus difficile dans des tâches trop ouvertes, comme être bon en mathématiques à l'école⁴⁷⁵.

Quels sont les facteurs qui prédisent mieux le niveau de maîtrise ? Pour des tâches qui ont un composant moteur important, la vitesse de progression au tout début est prédite par des capacités

⁴⁷³ Portion de la compétence qui est gérée de façon automatique, l'automatisation absolue étant la quantité totale de connaissance automatisée.

⁴⁷⁴ Ackerman, "New Developments in Understanding Skilled Performance."

⁴⁷⁵ Ibid.

générales (raisonnement spatial, intelligence verbale ou capacité de calcul). Dans la phase d'automatisation plus avancée, elle est prédite plutôt par des mesures de vitesse de perception et par des capacités psychomotrices. Sur des tâches plus intellectuelles, la distinction fondamentale est entre des tâches logiques, abstraites et dépourvues de contenu substantiel et des tâches avec un contenu substantiel. Dans le premier cas, des mesures de la capacité de raisonnement et de la mémoire de travail sont de bons indices de la performance ; dans le deuxième, le bagage des connaissances passées pertinentes est décisif. Dès qu'on abandonne les tâches abstraites dénuées de signification pratique, le plus important pour apprendre vite est d'avoir des connaissances pertinentes pour la nouvelle tâche. Cela va à l'encontre de l'idée selon laquelle les jeunes apprennent en général plus vite que les adultes. En ayant davantage d'expérience et de culture générale, souvent les adultes peuvent atteindre plus facilement un niveau de maîtrise dans les tâches intellectuelles. L'avantage des jeunes, comme nous le verrons, se situe plutôt au niveau de l'excellence, qui demande des véritables adaptations psychophysiques, pour lesquelles la plasticité des jeunes constitue sans doute un atout.

Jusqu'à il y a quelques décennies, la psychologie de l'excellence était considérée comme une extension de la psychologie de la maîtrise : une simple version extrême et plus poussée. Aujourd'hui, il paraît clair qu'il y a des différences psychologiques fondamentales entre les deux niveaux de performance. En ce qui concerne l'excellence, s'il y a plusieurs types de champions, les chemins pour y arriver se ressemblent énormément. Il a été montré que, pour arriver parmi les meilleurs 10% au monde, il faut au moins 10 ans non pas simplement de pratique ou d'entraînement en général, mais d'« entraînement délibéré » (*deliberate practice*)⁴⁷⁶. Pendant le processus d'acquisition de la maîtrise, la progression s'accompagne d'une automatisation accrue qui permet d'atteindre davantage de performance avec moins d'effort. Après la phase extrêmement dure du début, les choses vont de mieux en mieux. Mais ce processus plafonne à un niveau où pour progresser il faut recommencer un entraînement plus intensif, planifié et ciblé que celui du départ. Dans un entraînement délibéré, il faut être guidé par un enseignant de très haut niveau qui émet des diagnostics de la situation analytiques et précis, décelant à chaque fois ce qu'il faut travailler en priorité et par quelle stratégie. Sans ce travail systématique sur les points faibles, le niveau de performance plafonne. Il a été montré que, contrairement à l'opinion reçue, après une période de rodage de deux ou trois ans, la performance des médecins ne progresse pas avec l'âge et

⁴⁷⁶ K. A. Ericsson, "The influence of experience and deliberate practice on the development of superior expert performance," in *The Cambridge Handbook of Expertise and Expert Performance* (New York, NY: Cambridge UP, 2006).

l'expérience, sauf dans le domaine très restreint où ils ont eu l'occasion de pratiquer à très haut niveau et d'avoir beaucoup de *feedback*⁴⁷⁷.

Le bénéfice de l'entraînement délibéré est rigoureusement limité à la tâche spécifique : si on veut atteindre un niveau d'excellence dans une tâche qui paraît très proche, il faut refaire le processus ; il n'y a pas de transfert facile. Une capacité supérieure de mémorisation des positions des pièces ne se transmet même pas entre des jeux qui partagent les mêmes pièces et plateau. Les joueurs de GO ne sont pas en mesure de mémoriser rapidement une position de GOMOKO, et vice-versa⁴⁷⁸. Cela va à l'encontre de l'idéal d'excellence dans plusieurs domaines qui est propre à la Renaissance : ce qu'on peut avoir est plutôt une *maîtrise* dans plusieurs tâches, alors que l'excellence en deux domaines est déjà rare. Cela motive aussi des doutes sur l'étendue de l'excellence déclarée : en général, lorsqu'il n'y a pas de critère de succès net (gagner des matchs de tennis), on gonfle les compétences de ceux qui ont la réputation d'experts. Cela a été démontré au fil des décennies, lorsque des experts « nominés » affichaient des performances modestes dans des expériences contrôlées. Plusieurs études montrent que l'expérience et les qualifications des psychothérapeutes ne sont pas corrélées avec le résultat de leurs traitements⁴⁷⁹. L'excellence des médecins est rigoureusement circonscrite au traitement de certaines maladies : pour le reste, leur performance baisse souvent avec l'âge⁴⁸⁰.

Le scepticisme sur la nomination d'experts est aujourd'hui si répandu que la psychologie de l'expertise récente (deux dernières décennies) se base quasi uniquement sur l'excellence prouvable ou documentée. Il s'agit du paradigme dit de la performance de l'expert (*expert-performance*), où on cherche à isoler des tâches qui (1) peuvent être reproduites de façon systématique ; (2) demandent une réponse immédiate à une situation spécifique ; (3) sont censée caractériser la performance de l'expert : par exemple, choisir un coup à partir d'une certaine position aux échecs⁴⁸¹ ou, pour le tennis, simuler la réponse à un service dont la vidéo est projetée sur un grand écran. Le paradigme a été appliqué pour différencier le novice et l'expert dans plusieurs domaines, tels que le

⁴⁷⁷ Ericsson, "Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance."

⁴⁷⁸ M. Eisenstadt et Y. Kareev, "Aspects of human problem solving: The use of internal representations," *Explorations in cognition* (1975).

⁴⁷⁹ Ericsson, "The influence of experience and deliberate practice on the development of superior expert performance."

⁴⁸⁰ N. K. Choudhry, R. H. Fletcher et S. B. Soumerai, "Systematic Review: The Relationship between Clinical Experience and Quality of Health Care," *Annals of Internal Medicine* 142, no. 4 (2005).

⁴⁸¹ Première expérience qui a lancé le protocole, voir Adde Groot, *Thought and Choice in Chess* (Mouton, 1978).

diagnostic médical, la performance musicale, la danse, plusieurs sports, l'écriture, la peinture et la planification d'expériences⁴⁸².

S'il faut dix ans d'entraînement délibéré, il doit y avoir une raison psychologique profonde. Dans une première phase, vers le début des années 70, la recherche a convergé vers le modèle de l'excellence comme « optimisation » de l'utilisation de ressources cognitives standard. Dans ce cadre, on avait du mal à rendre compte des changements profonds dans la psychologie de l'expert⁴⁸³. L'hypothèse était la suivante : l'expert doit optimiser le fonctionnement cognitif à l'intérieur des mêmes limites très strictes d'attention, perception et mémoire du novice. En s'attaquant avant tout à la perception et à la mémorisation, la solution proposée se basait sur le concept de *chunking*. Etant donné qu'il y a des limites à la quantité de facteurs qui peuvent être remarqués et mémorisés, il faut que l'expert travaille sur des « unités cognitives » plus larges. Par le même processus qui nous permet de mémoriser une phrase mot par mot au lieu de lettre par lettre, le joueur d'échecs développe une sorte de vocabulaire pour mémoriser les positions dans l'échiquier, ce que lui évite de mémoriser directement les positions pièce par pièce.

Dans le jeu d'échecs comme dans les autres tâches, le choix des mots de ce « vocabulaire perceptif » est axé sur des principes stratégiques pertinents. Ainsi, le joueur voit une position défensive difficile ; un pompier expérimenté catégorise les situations de façon dynamique, par rapport au type d'incendie de départ et aux évolutions possibles (alors qu'un novice se concentre sur des aspects perceptifs plus superficiels)⁴⁸⁴. La perception de l'expert concentre une analyse stratégique dans un vocabulaire technique poussé et spécifique à la tâche. Lorsqu'une situation n'a aucun sens par rapport à cette lecture stratégique, par exemple un échiquier où les pièces sont placées au hasard, la capacité de mémorisation de l'expert retombe à des niveaux proches de ceux du novice.

Dans un premier temps, ces données semblaient étayer l'hypothèse de la simple optimisation, opérée sous strictes contraintes cognitives. Si on augmente la dimension et la spécialisation des unités cognitives, c'est justement parce que le nombre maximum d'unités traitables par le système est limité. C'est comme si, pour apprendre par cœur une phrase, on mémorisait mot par mot (expert)

⁴⁸²Nous verrons que cela a des conséquences intéressantes pour l'idéal du sage vertueux mis en avant par l'éthique de la vertu, notamment, par rapport au problème de l'identification très difficile des sages (voir chapitre sur la sagesse) K. A. Ericsson et P. Ward, "Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance," *Current Directions in Psychological Science* 16, no. 6 (2007): 348.

⁴⁸³ H. Chase et H. A. Simon, "Skill in chess," *American Scientist* 61, no. 4 (1973).

⁴⁸⁴ G. A. Klein, *Sources of Power: How People Make Decisions* (MIT Press, 1998).

au lieu de lettre par lettre (novice). Mais d'autres données ne sont pas compatibles avec l'hypothèse. Les champions d'échecs peuvent mémoriser des positions au-delà des limites standard pour la mémoire à court terme : ils peuvent se souvenir de positions précises dans des matches joués des mois auparavant. Le *chunking* est sans doute réel, mais il ne peut expliquer à lui seul ce phénomène : dans leur tâche spécifique, les experts montrent une capacité de mémorisation qui dépasse les contraintes cognitives standard de l'espèce. Il s'agit d'une véritable adaptation de la psychologie humaine à la tâche, permettant de stocker une quantité énorme de données dans la mémoire de travail à long terme⁴⁸⁵. La capacité à mobiliser un grand nombre de cas et situations intéressants permet d'avoir un grand répertoire pour confronter et évaluer rapidement la situation présente.

L'hypothèse de l'adaptation rend compte de façon beaucoup plus plausible du parcours long et difficile qui mène à l'excellence. Si optimiser l'exploitation du même stock de ressources cognitives était suffisant, il faudrait probablement moins de dix ans d'entraînement intensif et délibéré. Refaçonner le système cognitif pour une adaptation optimale à la tâche justifie l'effort nécessaire et explique l'avantage des jeunes dans la poursuite de l'excellence⁴⁸⁶ en termes de plasticité. La spécificité de l'adaptation explique aussi la difficulté du transfert, qui demanderait nécessairement une restructuration du système cognitif. L'entraînement délibéré est en mesure de produire des changements psychophysiques majeurs : cela a été démontré dans plusieurs domaines. Le physique d'un athlète de haut niveau est adapté de façon subtile à son activité ; un joueur d'échec a une mémoire prodigieuse des matchs passés. Il semble plausible que l'adaptation profonde active un profil de gènes spécifiques qui normalement ne s'exprimeraient pas, si bien que ceux qui atteignent un niveau d'excellence sont de véritables « mutants fonctionnels ». La question est débattue, mais il semble que la plupart des gènes nécessaires pour ces adaptations induites sont présents chez tout

⁴⁸⁵ La mémoire à long terme permet de stocker à long terme une sélection très restreinte de données significatives : par exemple, il est impossible de se souvenir des détails de toute conversation qu'on a eue, alors qu'on peut retenir le contenu des conversations les plus importantes. La mémoire de travail permet de stocker à très court terme un nombre restreint de détails (la phrase que vous venez de lire, un numéro de téléphone). Dans son domaine, l'expert développe une mémoire de travail à long terme qui permet de stocker à long termes le type de détails techniques qui seraient sans doute éliminés par la mémoire à long terme et qui ne seraient retenus que par quelques seconds par une mémoire de travail normale. K. A. Ericsson et W. Kintsch, "Long-Term Working Memory," *Psychological Review* 102, no. 2 (1995).

⁴⁸⁶ Mais non la maîtrise, au moins dans des tâches intellectuelles où la composante motrice est faible.

enfant sain. Les théoriciens les plus radicaux pensent que les seules limites au pouvoir de façonnement de l'entraînement sont les dimensions et la taille du corps⁴⁸⁷.

Sans prendre position dans ce débat sur le rôle résiduel joué par le talent, il est clair que la différence de dotation génétique se manifeste surtout au début de la courbe d'apprentissage, dans la difficulté à atteindre un niveau de maîtrise. Tout le monde a fait l'expérience de cela à l'école : par exemple, certains étudiants réussissent facilement en mathématique et d'autres peinent à avoir des résultats décents avec une quantité considérable d'effort. Mais en ce qui concerne le passage de la maîtrise à l'excellence, dix ans d'entraînement délibéré ne sont épargnés à personne (même s'il se confirme que les non-talentueux doivent payer un prix encore plus élevé). À l'état actuel de nos connaissances, le débat talent/entraînement ne soulève pas de doutes sur le fait suivant : l'excellence est rare, difficile à atteindre et tend à épuiser le potentiel de plasticité de l'organisme. En tout cas, elle est suffisamment difficile et coûteuse pour mettre à mal l'idéal du sage vertueux, chez qui toutes les formes d'excellence morale trouveraient leur accomplissement.

Compétence et vertu dans la tradition philosophique

Rien ne montre que la tradition philosophique ancienne sur la vertu ait conçu une différence significative entre la psychologie de la maîtrise et de l'excellence. Cela ne signifie pas que la distinction entre maîtrise et excellence n'était pas importante dans les théories de la vertu. Au contraire, s'approcher de l'idéal de la vertu a été conçu comme le travail d'une vie d'effort : le perfectionnement ne venait qu'avec l'âge. Mais cela n'était pas marqué par une discontinuité par rapport aux processus psychologiques qui produisent la compétence. L'excellence était conçue comme une forme de compétence pleine et parfaite. Il n'y avait pas de changement de modèle psychologique.

Le sujet de débat le plus important dans l'antiquité est celui qui garde encore aujourd'hui le plus grand intérêt (aussi bien en psychologie qu'en philosophie) : l'opposition entre une conception intellectualiste et une conception pratique de la compétence⁴⁸⁸. Le point de vue intellectualiste est représenté de façon claire et influente par Platon, qui l'attribue à Socrate. Une vraie compétence aurait trois propriétés : (1) il faut un processus explicite d'enseignement pour qu'elle soit apprise ; (2) celui qui acquiert vraiment une compétence doit comprendre les principes théoriques qui la

⁴⁸⁷ Ericsson, "Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance."

⁴⁸⁸ D. S. Hutchinson, "Doctrines of the mean and the debate concerning skills in fourth-century medicine, rhetoric and ethics," *Apeiron* 21, no. 2 (1988).

règlent ; (3) celui qui possède la maîtrise doit être en mesure de justifier son action dans le domaine pertinent⁴⁸⁹.

D'après le premier point, tout ce qu'on peut apprendre de manière satisfaisante grâce à l'imitation et à l'interaction avec des personnes compétentes ne constitue pas une vraie compétence, mais une espèce de talent naturel. Il faut qu'à la pratique spontanée et imitative s'ajoute l'enseignement théorique explicite d'un maître. Comme nous verrons dans la suite, pour évaluer une telle proposition il faut fixer une définition opérationnelle d'« enseignement théorique ». Bien sûr, si je commence un entraînement de tir à l'arc par une description de la physique pertinente et des limitations qu'elle impose, mon approche sera très théorique. Mais quel genre de correction de la part du maître commence à compter comme l'exposition d'une théorie ? Les philosophes n'ont pas un avis aussi tranché qu'il le faudrait sur la question. Par exemple, Annas cite l'exemple de l'apprentissage d'une langue étrangère comme activité qui demanderait nécessairement ce type d'apprentissage explicite :

« On n'apprend pas naturellement l'italien sans des enseignants, et sans effort de notre part. (Les gens qui apprennent l'italien sur le tas ont aussi bénéficié d'enseignants ; simplement, ils ne l'ont pas fait par le biais de cours, leçons de grammaires et choses de ce genre. Et il y a aussi des personnes qui habitent en Italie et n'ont jamais appris l'italien, car ils n'ont pas fait d'effort)⁴⁹⁰ »

Or, s'il faut bien sûr de l'effort pour apprendre une langue, ce genre d'exemple place très bas la barre de ce qu'on devrait considérer comme apprentissage théorique. Recevoir continuellement des corrections pertinentes est certainement plus que suffisant pour apprendre correctement l'italien, sans besoin d'introduire dans l'apprentissage les concepts de pronom, verbe, etc.

En général, il faut opérer une distinction entre au moins trois niveaux possibles de théorisation : (1) des corrections (ou exemples positifs) justifiées par le maître (« il ne faut pas tirer de cette façon-là, sinon tu vas casser l'arc, dont cette partie est la moins résistante ») ; (2) l'usage récurrent de véritables termes théoriques dans les corrections (« ceci est un pronom, tu ne peux pas l'utiliser comme cela ») ; (3) une véritable systématisation des propriétés de ces termes théoriques (une grammaire d'une langue, un traité de tir à l'arc). En réalité, il est clair qu'il s'agit d'un continuum : un maître pourrait aller au-delà de simples corrections justifiées, mais introduire seulement une poignée de termes théoriques ; il pourrait mobiliser un grand nombre de ces termes de façon plutôt cohérente, mais restant en deçà du niveau nécessaire pour un vrai traité.

⁴⁸⁹ Julia Annas, "Virtue as a skill," *International journal of philosophical studies* 3, no. 2 (1995), M. Stichter, "Ethical Expertise: The Skill Model of Virtue," *Ethical Theory and Moral Practice* 10, no. 2 (2007).

⁴⁹⁰ Annas, "Virtue as a skill," 372.

Une autre distinction pertinente est celle entre une visée purement pédagogique et une visée descriptive de la théorie proposée par le maître. Une théorie peut avoir une fonction purement descriptive, sans aucune utilité pédagogique : cela est le cas d'une description formelle extrêmement avancée, technique et détaillée de la grammaire d'une langue, ou une analyse des zones d'activation du cerveau pendant un processus réussi de persuasion. Mais l'intérêt purement scientifique d'une telle théorie ne nous concerne pas ici, car il est question de l'acquisition d'une compétence⁴⁹¹. La théorie doit avoir forcément une utilité pédagogique et elle n'est pas obligée de décrire correctement les processus psychologiques. Il y a des façons d'appréhender et catégoriser une situation pratique qui augmentent le succès sans décrire correctement le processus psychologique sous-jacent : simplement, interpréter la situation de cette manière aide à réussir. Plusieurs typologies humaines relèvent de ce cas de figure : catégoriser un client comme un têtard ou un rêveur peut aider pour réussir à le convaincre, même si une distinction aussi vague et simplificatrice a très peu de chance de décrire les processus causaux responsables de la réussite. Mais la théorie peut avoir aussi une ambition descriptive. Cette ambition restera toujours subordonnée à l'efficacité pédagogique : il n'est pas question d'introduire des détails techniques inutiles pour l'élève. Une distinction supplémentaire est celle entre le rôle pédagogique de la théorie pendant l'apprentissage et son rôle (éventuel) dans la pratique de l'expert. Par rapport à ce dernier, il y a deux hypothèses à évaluer : l'expert a besoin de réfléchir explicitement en mobilisant la théorie pendant son activité ; l'expert doit pouvoir expliquer et rendre compte *a posteriori* de ses choix, dans les termes de la théorie.

Avec ces distinctions à l'esprit, on peut voir que l'intellectualisme socratique défend le rôle d'une systématisation théorique plutôt importante, peut-être pas un vrai traité mais au moins des principes généraux qui pourraient en être la base. Une vraie compétence dans des tâches trop restreintes est exclue : on ne peut être expert d'Homère sans être expert de poésie en général. « *S'il existe quelque chose comme une maîtrise de la poésie, alors cela consiste en la compréhension des principes qui gèrent tout le domaine. Echouer dans une partie est échouer tout court*⁴⁹² ». Les compétences sont comme des espèces naturelles. On ne peut pas découper et recomposer des bouts ici et là : être expert uniquement de cinéma thaïlandais ou nager dans un seul style. La nage suit des principes et

⁴⁹¹ En réalité, une théorie descriptive détaillée peut avoir des effets pratiques très puissants mais indirects, par exemple la possibilité de manipuler le milieu d'apprentissage pour maximiser le résultat, grâce à la connaissance intime du processus psychologique pertinent. Nous nous occupons ici uniquement de l'effet direct de la théorie sur le novice.

⁴⁹² Annas, "Virtue as a skill," 231.

on doit les maîtriser tous pour maîtriser ne serait-ce qu'un seul style.⁴⁹³ Pour apprendre vraiment, il faut qu'un maître nous transmette ces principes dans l'apprentissage, et le maître lui-même doit être en mesure de justifier ses actions à la lumière de la théorie, si on le lui demande. Seule concession à l'évidence phénoménologique : il n'est pas nécessaire que l'expert réfléchisse explicitement sur ces principes pendant l'activité (rôle actif dans la pratique).

L'idée selon laquelle la théorie intervient dans tout apprentissage pratique paraît contraire à l'expérience commune : c'était déjà le cas au quatrième siècle avant notre ère, chez Isocrate et d'autres rhétoriciens, défenseurs d'une conception pratique ou empiriste de l'expertise. Isocrate soulignait l'importance du talent naturel, la pratique, l'entraînement et l'apprentissage avec un maître. Ce modèle, repris largement par Aristote, ne se différencie de l'approche intellectualiste que par un degré inférieur de systématisation dans la théorie apprise. Cette dernière n'est pas éliminée totalement, mais on considère qu'il s'agit de l'usage de termes théoriques peu ou pas systématisés, de règles générales de conduite vagues et qui admettent beaucoup d'exceptions. Il ne s'agit pas d'une évacuation totale : cela ressort du traitement très représentatif du statut de la médecine. D'après les défenseurs de la position intellectualiste, tels que Hippocrate, l'expérience ne peut que manifester des régularités approximatives : si on se limitait à cela, la pratique médicale serait abandonnée à l'imprécision et à la non-fiabilité. Pour retrouver la précision, il faut s'inspirer d'un corpus de connaissances plus larges, dont des faits généraux sur la nature de l'univers. Au contraire, d'après les empiristes : « *la médecine a avancé jusqu'à présent simplement grâce à l'expérience, l'essai et l'erreur, sans théories ou hypothèses profondes (...) après une période suffisamment longue d'adhésion à cette méthode non théorique et, pourrait-on dire, empirique, la médecine trouvera son accomplissement*⁴⁹⁴ ». L'exemple de la médecine montre que les empiristes admettent un certain niveau d'élaboration théorique. La cible de leur critique est la systématisation, non la réflexion en tant que telle. En d'autres termes, ils ne défendent pas un modèle radicalement intuitionniste de l'expertise.

Par contre, les théoriciens contemporains de la vertu, qui reprennent le débat ancien portant sur la codifiabilité de la connaissance pratique, insistent sur le rôle de l'intuition de l'expert, par

⁴⁹³ La psychologie contemporaine nous enseigne que l'expertise est fragmentée, c'est le sujet de ce chapitre. Toutefois, il y a peut-être une limite à la fragmentation: on ne peut pas être un expert mondial uniquement d'une échelle musicale spécifique (par exemple, être capable de la jouer dans n'importe quel contexte à une vitesse impressionnante), tout en étant parfaitement médiocre dans tout le reste. Ce serait intéressant que la psychologie se pose la question de l'extension minimale possible d'un domaine d'expertise.

⁴⁹⁴ Hutchinson, "Doctrines of the mean and the debate concerning skills in fourth-century medicine, rhetoric and ethics," 27.

opposition à la réflexion⁴⁹⁵. Le niveau de réflexivité et le niveau de codification sont deux questions distinctes, car la réflexion n'entraîne pas le raffinement théorique. Du point de vue de l'adéquation empirique de la théorie de la vertu, l'intérêt d'une conception fortement intuitionniste est évident : (1) la vertu est difficile à apprendre ; (2) tout ce qui relève entièrement d'une perception intuitive ou d'automatismes devient facile, libérant des ressources cognitives pour le reste, donc, (3) la vertu a intérêt à être aussi intuitive que possible. La stratégie des défenseurs d'une conception fortement intuitionniste de la vertu est ingénieuse et peut compter sur le soutien inattendu de plusieurs résultats de la recherche contemporaine sur la cognition automatique.

Schématiquement, la psychologie de la prise de décision a trois phases : interprétation de la situation, délibération sur quoi faire et comment et, en cas de conflit entre deux ou plusieurs stratégies d'intervention, résolution de ce conflit et prise de décision finale. Pour réduire le rôle de la réflexion explicite dans l'exercice d'une compétence, il y a donc trois leviers : (1) augmenter le plus possible la puissance de la perception pratique directe, si bien qu'un expert pourra très souvent « voir » la structure d'un problème pratique (au lieu de réfléchir sur l'interprétation correcte de la situation) ; (2) automatiser la délibération ; (3) soutenir qu'il y a des processus automatiques et pré-réflexifs de gestion de conflits méta-cognitifs. Les trois stratégies se soutiennent et se compensent, car chacune libère de la charge cognitive, rendant plus plausible la conception intuitionniste de la compétence. Le théoricien intuitionniste de la vertu trouve un large soutien dans la psychologie contemporaine. Plusieurs littératures psychologiques portant sur le pouvoir de l'intuition (notamment dans la cognition morale) et sur la cognition automatique et non réflexive convergent vers une réduction spectaculaire du rôle de la réflexion explicite.

La psychologie de la compétence d'après Dreyfus

Le modèle intuitionniste par excellence, et le seul qui soit cité par les philosophes (à ma connaissance), est celui proposé par Dreyfus dans le cadre de son approche phénoménologique. Dreyfus repère cinq phases dans l'apprentissage : novice (*beginner*), novice avancé (*advanced beginner*), confirmé (*competence*), maîtrise (*proficiency*), excellence (*mastery*)⁴⁹⁶. Le novice

⁴⁹⁵ Le représentant le plus extrême d'une approche intuitionniste à la psychologie de la vertu est John McDowell. McDowell, "Virtue and Reason."

⁴⁹⁶ S. E. Dreyfus, "The Five-Stage Model of Adult Skill Acquisition," *Bulletin of Science, Technology & Society* 24, no. 3 (2004). Dans certains textes, Dreyfus ajoute une phase supplémentaire, définie par le fait que l'expert développe son propre style. Pour atteindre cette phase, il faut avoir étudié avec plusieurs maîtres. Si on apprend d'un seul

commence par repérer des facteurs saillants accessibles à une perception ordinaire, et leur associe des règles d'action explicites. Par exemple, il va apprendre la fonction des pédales, du volant et de la boîte de vitesse dans une voiture, ainsi que la procédure schématique pour faire partir le véhicule. Dans une deuxième phase, la perception du novice (désormais avancée) commence à s'enrichir et il voit des choses qui ne peuvent pas être reconnues par un débutant absolu. Par exemple, il se rend compte que le bruit du moteur est anormal ou que le roi est peu défendu dans une position d'échec. À ce point, le novice avancé a encore besoin de règles, mais il peut les appliquer à un nombre plus important de situations, grâce au nouveau « vocabulaire perceptif » acquis. Plus précisément, si on retient la distinction entre règle (applicable par tout le monde, comme dans un manuel d'instructions) et maxime ou précepte (applicable uniquement par quelqu'un qui a déjà un certain niveau de compétence)⁴⁹⁷, on peut dire que le novice avancé commence à utiliser des maximes.

La puissance perceptive de l'élève confirmé est beaucoup plus importante : il remarque une quantité considérable de détails pertinents et il commence à rencontrer du succès dans la pratique. L'obstacle plus grand est constitué par des doutes sur la façon d'interpréter certaines situations. L'élève confirmé fait des analyses pertinentes, mais il hésite sur la perspective interprétative à adopter : aux échecs, il peut voir qu'il faut profiter d'une faute de l'adversaire, mais il hésite entre une attaque décidée et une stratégie défensive qui ne affaiblit pas sa défense. Il voit les options, mais il hésite dans la modulation de la réponse. D'après Dreyfus, c'est aussi la phase où de l'engagement émotionnel commence à se développer : du plaisir dans le succès et de la déception dans l'échec. Le débutant se concentre sur la technicité de la tâche et reste plutôt froid : sans engagement, l'élève confirmé peine à avancer.

Au niveau de la compétence, l'interprétation de la situation est entièrement automatisée et le but est clair : il ne reste qu'à réfléchir aux détails de l'implémentation. Un chauffeur expérimenté doit réfléchir pour négocier un tournant difficile (freiner ou lever l'accélérateur) ; un maître d'échec a besoin de peaufiner les détails de son attaque. Dans cette phase, les règles et les maximes vont être progressivement abandonnées : les problèmes qui restent sont techniques et relèvent d'une modulation des moyens, non pas de stratégies et orientations. Si on prend plaisir à l'exercice de sa compétence, l'intériorisation est telle que l'on commence à pouvoir résoudre les problèmes de façon froide et détachée.

maître, on finit par reproduire à l'identique le style du maître, sans aucune originalité. Cet ajout ne change en rien la position de Dreyfus sur l'automatisation. Voir H. L. Dreyfus, *On the internet* (Routledge, 2001).

⁴⁹⁷ Par exemple, aux échecs : « ne laisse pas le roi sans défense adéquate ». Le novice absolu n'est pas en mesure de juger ce qui compte comme « défense adéquate ».

La psychologie du niveau d'excellence est vite résumée : tout est automatique. « *Ce qui doit être fait, est fait* ». pourtant, pas de discontinuité psychologique par rapport à la maîtrise : « *celui qui maîtrise n'a simplement pas eu suffisamment d'expérience, à cause de la grande variété de réponses possibles au nombre de situations qu'il peut discriminer* » alors que « *avec suffisamment d'expérience dans une variété de situations, toutes vues depuis la même perspective mais demandant des décisions tactiques différentes, le cerveau de l'expert décompose graduellement cette classe de situations en sous-classes, dont chacune demande une réponse spécifique. Cela permet la réponse situationnelle immédiate qui est typique de l'expertise*⁴⁹⁸. » La continuité psychologique est totale : l'expérience est simplement le résultat d'une quantité énorme de délibérations sur les détails d'implémentation (typiques de la phase de maîtrise), jusqu'au point où il n'y a plus de détails qui nous surprennent. On retrouve ici la même conception de la continuité entre maîtrise et excellence qui caractérise la théorie de la vertu ancienne : l'excellence est atteinte progressivement et naturellement, après des années d'efforts et d'expérience.

D'un point de vue psychologique, le point le plus solide de cette reconstruction concerne le rôle de la perception. Par contre, la toute-puissance des automatismes (aux dépens de la réflexion) n'est pas défendable, en ce qui concerne la phase de l'excellence.

Evaluation du modèle intuitionniste : le rôle de la réflexion

Personne ne conteste l'importance de l'automatisation dans l'acquisition de l'expertise. Pourtant, les ambitions des intuitionnistes ne vont pas de soi : l'exercice d'une compétence se présente comme un mélange d'habitude et de délibération difficiles à démêler. L'habitude est essentielle, mais la réflexion ou l'attention explicitement dirigée ne gardent-elles pas des rôles clés ? Face à l'argument phénoménologique, le défenseur d'une conception plus délibérative de l'excellence pratique avance trois réserves : (1) *initiation de l'action* : l'attention explicite et une forme (aussi minimale soit-elle) de délibération gardent un rôle dans l'initiation du cours d'action (prise de décision, « quoi faire ? ») ; (2) *complexité* des processus automatisés : l'attention réflexive garde une place de directeur d'orchestre dans l'activation d'une kyrielle de petits processus et modules automatisés, car ce qu'on peut automatiser complètement doit être simple d'un point de vue cognitif ; (3) *portée* des processus automatisés : la réflexion doit forcément jouer un rôle dans la résolution des conflits entre plusieurs modules automatisés, car il n'est pas possible d'automatiser « en bloc » et de façon harmonieuse une compétence trop large (ex : si un pilote professionnel peut

⁴⁹⁸ Dreyfus, "The Five-Stage Model of Adult Skill Acquisition," 179-180.

automatiser sa capacité à dépasser en piste, une automatisation complète de « conduire une voiture de formule 1 », avec tout ce que cela comporte, n'est pas possible). La littérature des dernières décennies sur la cognition automatique répond aux deux premières formes de scepticisme, mais elle est moins convaincante sur la troisième.

La notion d'automatisme en psychologie

L'étude de la cognition automatique, dont Jonathan Bargh est l'un des principaux pionniers, se situe dans le contexte des « théories duales » de la cognition. Selon cette approche, il y aurait un système cognitif plus ancien d'un point de vue évolutionniste et caractérisé par des processus (1) dont on n'est pas conscient, (2) autonomes (capables de se dérouler sans intervention dès qu'ils se sont déclenchés), (3) non volontaires (non déclenchés par un acte volontaire de l'agent) et (4) qui ne demandent pas d'effort. À ce système, l'évolution aurait ajouté un système de traitement de l'information qui (1) produit de l'action intentionnelle, (2) mobilise des processus dont on est bien conscient, (3) est flexible parce que sujet à contrôle et (4) qui demande de l'effort cognitif⁴⁹⁹.

Dans une première phase de la recherche (milieu des années 70), on avait tendance à penser que tout processus était soit parfaitement automatique soit parfaitement contrôlé, sans une zone grise entre les deux. Plus récemment, on s'est rendu compte que les différentes caractéristiques peuvent figurer dans plusieurs combinaisons. Il est possible de catégoriser les différents types de processus automatiques par « contexte de déclenchement » et « profil d'automatisme ».

Contexte de déclenchement

Les conditions dans lesquelles un processus se déclenche peuvent être catégorisées de la façon suivante. D'abord, tout contexte d'activation est formé par le stimulus externe (contexte d'activation externe) et la condition psychologique interne requise (contexte d'activation interne). Dans des cas extrêmes, l'un domine complètement l'autre. Un maniaque est sous le contrôle total d'un stimulus ; il ne peut s'empêcher de réagir d'une certaine façon lorsqu'il le perçoit, quelle que soit sa condition mentale. Inversement, il est possible d'activer une attitude constamment, quel que soit l'environnement. Normalement, on retrouve toujours un composant interne et un composant externe : dans la psychologie d'un individu, un processus automatique se déclenche à partir d'un stimulus et d'un état mental spécifiques.

⁴⁹⁹ J. A. Bargh, "Conditional Automaticity: Varieties of Automatic Influence in Social Perception and Cognition," in *Unintended Thought*, édition J. A. Bargh et J. Uleman (New York: Guilford, 1989), 3-4.

En ce qui concerne le contexte de déclenchement externe, suivant que l'agent est conscient ou non du stimulus d'activation, on parlera d'automaticité post-consciente ou pré-consciente : je ne peux pas résister aux gâteaux, et bien évidemment je suis conscient de les voir lorsque je décide de manger (post-conscient) ; je ne suis pas conscient de mes réactions très précises à la gestuelle de mon partenaire (pré-conscient). En ce qui concerne le contexte de déclenchement interne, si le déclenchement du processus requiert la présence d'un but ou d'une motivation préalables, on parlera d'automaticité dépendante d'un but (*goal-dependent*). La présence du but peut être évidente ou plus subtile. Par exemple, même si je me retrouve dans un court de tennis, je ne vais servir que si je veux jouer au tennis (but évident). Par contre, la formation automatique d'impressions est souvent influencée de façon inconsciente par des intérêts liés à la gestion de sa réputation : il s'agit d'une découverte informative.

Si le but qui est atteint dans les faits n'est pas celui qui était visé (et dont l'existence est nécessaire pour l'activation du processus automatique), on parle de processus non-intentionnel dépendant d'un but. Par exemple, lorsque j'ai envie de lire un livre, je finis toujours par surfer sur Internet. Plus généralement, tout processus qui se déclenche en tant qu'effet secondaire non cherché d'une autre activité relève de ce cas de figure.

Profil d'automaticité

Les variables du profil d'automaticité sont les niveaux de (1) contrôle, (2) présence à la conscience pendant l'exécution et (3) effort. Des niveaux élevés de ces variables ne co-occurrent pas nécessairement : une activité compulsive peut être initiée intentionnellement (ou au moins en pleine conscience) et après échapper au contrôle, tout en étant bien présente à l'attention de l'agent (ex. fumer une cigarette). Une action peut être initiée intentionnellement (je prends la voiture) et continuer après entièrement « en pilote automatique » (je conduis en réfléchissant à autres chose), tout en restant très facile à contrôler (je peux m'arrêter à tout moment). Un processus peut être initié de façon non intentionnelle (je commence à m'énerver et attaquer mon interlocuteur, car je suis jaloux) et redevenir accessible à l'attention et au contrôle après (je me rends compte de ce que je suis en train de faire et je me calme).

Le déclenchement automatique de l'action vertueuse

À la lumière de ces catégories, il semble que l'excellence dans une activité pratique ne se caractérise pas par un niveau maximal d'automaticité. Jouer au tennis n'est pas seulement une activité dépendant d'un but, elle est aussi initiée de façon intentionnelle (on ne peut pas commencer

à jour au tennis par hasard), sous contrôle (on peut jeter la raquette et arrêter de jouer quand on veut) et présente à l'attention (si on exécute une volée, un service ou un smash, on s'en rend compte). En plus, seules des séquences de mouvements très courtes sont autonomes (une fois déclenchées, elles se déroulent jusqu'au but sans intervention et attention). Par exemple, dès qu'un joueur expérimenté a l'intention de servir, il le fait sans penser aux phases intermédiaires du mouvement. Mais après, même sans délibérer, il faut au moins réinitialiser l'action : le joueur se rend compte qu'il est à l'origine de chaque coup qu'il enchaîne après le service. Dans cette reconstruction, l'exercice d'une expertise pratique ressemble à une activité « non explicitement délibérée » (par rapport à quoi faire et comment), mais segmentée et plutôt contrôlée d'après les standards des psychologues, qui ont appelé parfois ce cas de figure « actions semi-automatiques »⁵⁰⁰.

Le théoricien intuitionniste de la vertu vaudrait plus que cela, et il a raison. L'une des limites de l'analogie entre la maîtrise dans un sport (ou une discipline intellectuelle) et la vertu est que le champ d'action de la vertu est beaucoup plus général et moins lié à un contexte précis. On ne peut pas commencer à jouer aux échecs sans le vouloir. Par contre, les situations moralement pertinentes sont répandues et se fondent dans le tissu de la vie de tous les jours. On ne doit pas forcément se rendre compte qu'on commence à « jouer le jeu de la morale ». Dans un contexte d'action fortement chargé au niveau moral (la défense d'une ville contre des assaillants), on n'est pas obligé de remarquer le déclenchement d'une séquence d'action courageuse. Plus exactement – étant donné qu'il est difficile de courir vers l'ennemi ou d'ouvrir le feu sans se rendre compte qu'on est en train de le faire – il n'est pas nécessaire d'en être conscients sous la description de l'action pertinente pour l'excellence pratique et la vertu (en tant qu'action courageuse)⁵⁰¹. On n'a pas besoin de se dire « là je commence à être courageux ».

Les données psychologiques relatives à l'automaticité pré-consciente nous montrent qu'un déclenchement complètement automatique de la vertu est possible. Un exemple clair d'automaticité pré-consciente est l'activation de stéréotypes sociaux. Un stéréotype social s'active directement à partir d'un stimulus ou une situation pertinents et affecte notre réaction de façon autonome, sans besoin de contrôle ou d'effort pour que les différentes étapes du traitement cognitif du stéréotype se succèdent. Par exemple, lors d'un entretien avec une personne appartenant à une ethnie associée à

⁵⁰⁰ Ibid.

⁵⁰¹ Par contre, on ne peut pas ne pas se rendre compte qu'on commence à jouer au tennis ou à servir sous la description pertinente dans le jeu du tennis.

des connotations négatives, les biais d'évaluation subtils des déclarations du candidat s'enchaînent sans que la personne qui pose les questions s'en rende compte.

D'un point de vue cognitif, l'automatisme pré-conscient s'explique par l'activation directe d'une représentation cognitive (dans l'exemple, le stéréotype racial) par un stimulus externe, sans besoin de conditions facilitatrices telles qu'avoir réfléchi à la question, être dans un certain état d'esprit ou avoir un but spécifique. En psychologie socio-cognitive contemporaine, on parle dans ce cas d'accessibilité chronique. Une structure cognitive chroniquement accessible est une représentation qui est mobilisée très facilement. Dans ces littératures empiriques, on travaille plus ou moins explicitement avec un modèle connexionniste des rapports entre représentations cognitives et stimuli externes (voir chapitre sur les modèles socio-cognitifs). Le simple fait d'être à une fête (stimulus externe) active de façon plus ou moins intense les concepts de bière, amusement ou drague et inhibe l'idée d'étude ou travail. Indépendamment de la façon dont elle a été activée, l'idée de fête affectera le réseau d'activations du système cognitif.

Une représentation chroniquement accessible est activée très fréquemment, par un biais ou par un autre. Les concepts qu'on a de soi-même sont un bon exemple : si on pense être très timide, tout sera interprété à partir de cela. Les situations sociales seront classées comme plus ou moins dangereuses par rapport à la timidité : même des idées qui ont très peu à voir avec ce sujet seront saillantes si elles nous aident à justifier notre image de timide. L'association constante d'un stimulus à une représentation chroniquement accessible permet un déclenchement automatique pré-conscient. Il n'est pas possible d'être conscient des nombreuses distorsions du traitement de l'information sociale générées par une représentation chroniquement accessible.

La complexité des processus cognitifs automatiques

Le parallèle entre l'automatisme pré-conscient et l'exercice de l'excellence pratique est évident. Une séquence d'actions courageuses est déclenchée par une situation pertinente sans que l'agent se rende compte qu'il commence, procède et termine d'agir de façon courageuse⁵⁰². Toutefois, ici il faut répondre à la deuxième forme de scepticisme du défenseur d'une conception délibérative de l'expertise (l'idée selon laquelle on ne pourrait automatiser que des processus simples et mécaniques). Ce qu'on peut automatiser de façon aussi profonde n'est-il pas forcément simple d'un point de vue cognitif ? Notre exemple a été un stéréotype racial, alors que l'ambition était de

⁵⁰² Le parallèle est développé dans les détails par Nancy Snow, "Habitual Virtuous Actions and Automaticity," *Ethical Theory and Moral Practice*.

montrer la possibilité d'une compétence pratique complexe et complètement automatique. La richesse des effets en aval – insiste le sceptique – n'est pas une preuve de la complexité de la représentation activée en amont. Par exemple, entrer dans un mode « défensif » de justification de mon comportement en tant que timide peut affecter de façon subtile le traitement de l'information sociale, à plusieurs niveaux. Mais tout cela est déclenché de façon indirecte par la simple activation d'un trait. L'automatisation pré-consciente de dynamiques « intelligentes » et flexibles comme le courage est autre chose.

Heureusement, la recherche sur la cognition automatique a récemment prouvé que l'activation pré-consciente de dynamiques intelligentes est possible. Parmi les structures cognitives chroniquement accessibles, on ne trouve pas uniquement des stéréotypes, mais aussi des objectifs et des attitudes liés aux valeurs de l'agent, par exemple l'objectif de l'équité dans les interactions sociales, l'objectif d'être un bon parent, une personne qui a du succès, une personne moralement digne ou coopérative⁵⁰³. Les objectifs concernés peuvent aussi relever de styles cognitifs dans le traitement de l'information, tels que la volonté de comparer plusieurs alternatives ou de se concentrer sur les détails d'implémentation⁵⁰⁴.

Les domaines impliqués sont sans doute les bons, pertinents aussi bien pour les vertus morales qu'intellectuelles : il s'agit maintenant de voir si ces buts pré-conscients ont la complexité et la flexibilité requises.

Il faut trouver des critères pour distinguer l'activation de buts et stratégies de haut niveau de l'activation de représentations plus simples (traits, stéréotypes). Un aspect fondamental de l'activation d'un stéréotype ou d'un trait est la réduction rapide de son influence au fil du temps. Au début, l'effet peut être impressionnant : le simple fait de mentionner de façon saillante un trait conditionne l'interprétation des stimuli ambigus. Si vous venez de lire un texte sur la timidité, une action qui peut être interprétée de plusieurs façons vous paraîtra une manifestation de timidité. Au niveau comportemental, si la catégorie de timidité vient d'être activée, on devient un peu plus

⁵⁰³ J. A. Bargh, "Auto-Motives: Preconscious Determinants of Social Interaction," in *Handbook of Motivation and Cognition: Foundations of Social Behavior*, édition E. Higgins et R. Sorrentino (New York: The Guilford Press, 1990).

⁵⁰⁴ P. M. Gollwitzer, H. Heckhausen et B. Steller, "Deliberative and implemental mind-sets: Cognitive tuning toward congruous thoughts and information," *Journal of Personality and Social Psychology* 59, no. 6 (1990).

timide. Mais cela ne dure pas longtemps⁵⁰⁵. Par contre, un but qui n'a pas encore été atteint a tendance à devenir de plus en plus saillant au fil du temps⁵⁰⁶.

Etant donné que les effets de représentations simples et de haut niveau se superposent souvent⁵⁰⁷, il est essentiel de les dissocier pour prouver qu'il y a eu une véritable activation de la représentation implicite d'un but. Bargh a conduit une expérience dont les variables étaient les suivantes : (1) certains participants devaient essayer de former le plus grand nombre possible de mots à partir d'un ensemble de pièces marquées chacune par une lettre. D'autres étaient confrontés à une tâche de formation d'impressions : juger l'attitude d'une personne qui se comportait de façon ambiguë par rapport à la motivation de réussir ; (2) auparavant, certains participants avaient été exposés (ou non) à des mots liés au succès et la réussite dans une tâche de recherche de mots (dans une matrice de n x n lettres) ; (3) certains devaient exécuter les deux tâches l'une après l'autre, d'autres attendaient cinq minutes. Au total, la combinaison de ces trois variables crée huit sous-groupes (2x2x2).

Le résultat a montré une dissociation claire : après cinq minutes, chez ceux qui avaient été exposés aux mots liés à la réussite, il y avait une tendance beaucoup plus faible à interpréter le comportement ambigu comme une manifestation du désir de réussir (dans la tâche de formation des impressions), mais un succès beaucoup plus net dans la tâche de formation de mots. Cela montre qu'il y avait eu une activation implicite aussi bien de la catégorie de « réussite » (bas niveau) que du but de réussir (haut niveau), les deux phénomènes étant dissociables et associés à une évolution temporelle différente. L'affaiblissement avec le temps est lié à toute activation de catégorie perceptive⁵⁰⁸ : par conséquent, cette expérience permet d'écarter d'autres hypothèses sceptiques quant à l'activation d'un but. Par exemple, on peut écarter l'hypothèse selon laquelle il y aurait une perception différente de la situation chez ceux qui ont été exposés aux mots de réussite, si bien que la tâche de formation de mots serait perçue en elle-même comme étant plus urgente (« formez le maximum de mots, coûte que coûte »). Ce genre d'effet – de nature perceptive – devrait s'estomper très rapidement.

⁵⁰⁵ En général, aussi peu que cinq minutes, voir J. A. Bargh, W. J. Lombardi et E. T. Higgins, "Automaticity of chronically accessible constructs in person x situation effects on person perception: it's just a matter of time," *Journal of Personality and Social Psychology* 55, no. 4 (1988).

⁵⁰⁶ J. W. Atkinson et D. Birch, "A dynamic theory of action" (New York: Wiley, 1970), Bargh, Lombardi et Higgins, "Automaticity of chronically accessible constructs in person x situation effects on person perception: it's just a matter of time."

⁵⁰⁷ Une activation implicite du but d'être coopératif active probablement aussi « coopératif » en tant que simple trait.

⁵⁰⁸ Bargh, Lombardi et Higgins, "Automaticity of chronically accessible constructs in person x situation effects on person perception: it's just a matter of time."

Si on ne peut pas réduire l'automatisme pré-conscient à l'activation d'un trait ou stéréotype, encore faut-il montrer qu'elle a des caractéristiques normalement associées aux objectifs poursuivis de façon consciente, notamment la persistance face à des obstacles et la reprise de l'activité (visant le but) après une interruption. Dans une autre expérience de Bargh, les participants, après avoir été exposés (ou non) à des mots de réussite dans une tâche de recherche de mots, étaient interrompus pendant une deuxième tâche du même genre, deux minutes après le début⁵⁰⁹. Leur réaction face à cette interruption inattendue était filmée : seulement ceux qui avaient été exposés aux mots de réussite avaient tendance à continuer à travailler après l'ordre d'interruption donné par l'expérimentateur. Cela est une preuve de persistance face aux obstacles. Dans une dernière expérience semblable, au début de la deuxième tâche, les participants étaient interrompus tout de suite, sous prétexte d'un problème technique. Il s'agissait de former le plus possible de mots à partir de suites de lettres affichées avec un rétroprojecteur, dont on simulait un problème technique. On leur avait annoncé trois tâches différentes dans la session : suite au faux problème technique, après cinq minutes, les participants devaient choisir entre reprendre la même tâche ou commencer la troisième, censée être beaucoup plus amusante (juger si des dessins animés étaient drôles). Seuls les participants stimulés par des mots de réussite dans la première tâche tendaient à insister pour continuer avec la deuxième tâche, au lieu de s'amuser avec la troisième.

Globalement, ces données soutiennent de façon nette l'identification de l'action vertueuse avec l'automatisme pré-conscient, dans le cadre d'une approche intuitionniste à la vertu. Dans les expériences de Bargh, les structures cognitives activées sont les attitudes et valeurs profondes de l'individu (réussite), ainsi que son style cognitif, ce qui couvre le domaine typique des vertus morales et intellectuelles. Ces structures ont la complexité et flexibilité cognitives requises : il ne s'agit ni de la simple activation d'une catégorie perceptive (traits, stéréotypes), ni d'un simple réflexe comportemental contrôlé par un stimulus. Un réflexe de ce dernier type est impliqué dans la réaction à une tentation : on voit du chocolat et on veut le manger. Au contraire, l'automatisme pré-conscient est souvent recruté pour résister à de tels réflexes. Par exemple, la vision du chocolat peut activer en même temps la tentation de le manger et l'objectif de perdre du poids, ce qui nous permet de résister à la tentation⁵¹⁰.

⁵⁰⁹ Deux minutes et pas plus, parce que ceux qui étaient motivés à réussir n'avaient pas le temps nécessaire pour croire avoir déjà réussi. Pas moins, parce que tout le monde soit suffisamment engagé avec la tâche.

⁵¹⁰ A. Fishbach, R. S. Friedman et W. Arie, "Kruglanski (2003), "Leading Us Not into Temptation: Momentary Allurements Elicit Overriding Goal Activation,"" *Journal of Personality and Social Psychology* 84, no. 2, Snow, "Habitual Virtuous Actions and Automaticity."

Les limites de l'automatisation : le rôle de la réflexion

Il nous reste à traiter le scepticisme quant à la portée de l'automatisme pré-consciente : si des processus cognitifs de haut niveau peuvent être complètement automatiques, encore faut-il démontrer qu'on peut automatiser de la sorte un ensemble complexe et cohérent de compétences, y compris l'activité méta-cognitive nécessaire pour la coordination et la gestion des conflits – jusqu'au point où cela devienne une véritable deuxième nature. Cela soulève le problème des liens entre méta-cognition et conscience, attention et réflexion explicite. De premier abord, l'intuitionniste semble être dans une bonne position pour l'emporter aussi sur ce terrain, car la notion de méta-cognition n'est pas nécessairement liée à celle d'attention et réflexivité⁵¹¹.

Il y a maintenant des données claires sur l'existence de monitoring et contrôle méta-cognitifs complètement inconscients, dans le cadre de la littérature sur le choix des stratégies dans la résolution des problèmes. Cette littérature montre que les gens trouvent (ou apprennent) des stratégies de résolution et les adaptent suivant la situation et le taux de succès dans la tâche. Cette adaptation démontre en même temps le monitoring (le système cognitif est sensible au taux de succès de ses propres stratégies) et le contrôle (adoption d'une stratégie plus performante dans une situation donnée), tout cela sans que les gens se rendent compte ni de quelles stratégies suivent, ni qu'ils les adaptent et pourquoi. Voyons quelques exemples.

Reder et Schunn ont élaboré un problème de gestion du trafic aérien où il s'agissait de décider quels avions pouvaient atterrir dans quelles pistes d'atterrissage. Il y avait des avions grands (boeing 747) ou plus petit, et des pistes plus longues ou plus courtes. Les Boeing 747 a besoin d'une piste longue, pour les autres cela est indifférent. La variable cruciale est le pourcentage de 747 sur l'ensemble de tous les avions. Les deux stratégies optimales dans ce problème plutôt simple sont les suivantes : si le pourcentage de 747 est élevé, il vaut mieux ne pas faire atterrir les petits avions dans les pistes longues ; si le pourcentage est réduit, cette interdiction n'est pas justifiée. Au fil de plusieurs séances de ce jeu, les participants alternaient de façon correcte ces deux stratégies, sauf qu'ils n'en étaient pas conscients. Loin de se rendre compte des stratégies adoptées, ils ne réalisaient même pas

⁵¹¹ Voir T. O. Nelson et L. Narens, "Metamemory: A theoretical framework and new findings," *The psychology of learning and motivation* 26 (1990). Nous sommes largement débiteurs du traitement de la question dans Chris Lepock, "Metacognition and epistemic virtue" (thèse de doctorat: University of Alberta, 2007).

que le pourcentage d'avions 747 variait de façon sensible pendant le jeu. En d'autres termes, il n'y avait même pas conscience de la véritable structure du problème⁵¹².

Dans une autre expérience, il s'agissait de juger de la plausibilité d'affirmations relatives à une histoire qu'on venait de lire. Les affirmations pouvaient être soit la répétition (ou la simple négation) de quelque chose qui était explicitement dit dans l'histoire, soit des ajouts plausibles (ou non plausibles). On avait à disposition deux stratégies de réponse : un vrai jugement de plausibilité ou une récupération directe de la mémoire. Des délais de réponse plus rapides étaient censés être un signe de la dernière stratégie. Comme dans l'expérience précédente, le pourcentage des répétitions et des ajouts était varié au fil des essais. Même si les instructions explicites données par les expérimentateurs allaient dans le sens de l'une des stratégies (évaluer la plausibilité), lorsque la plupart des affirmations étaient des citations du texte, les participants adoptaient très rapidement une stratégie de récupération directe, à nouveau abandonnée si la situation changeait au fil de la session. Dans ce cas non plus, aucune conscience des stratégies (curieusement, les participants croyaient avoir toujours adopté une stratégie de récupération) ni de la structure du problème (variation du rapport répétitions/ajouts)⁵¹³.

Un effet semblable a été trouvé dans une tâche arithmétique de vérification de multiplication : sans faire aucun calcul, il est possible de remarquer que certaines multiplications sont fausses parce que le résultat ne peut pas être impair ($8 \times 7 = 57$). Lorsque le pourcentage de fausses multiplications de ce genre était élevé, les participants montraient des signes (temps de réactions beaucoup plus rapides) d'usage massif de cette stratégie, sans aucune conscience du phénomène. D'autres adaptations stratégiques inconscientes sont moins frappantes : notre choix stratégique est souvent sensible à la disponibilité d'aides pour la mémoire de travail (papier, stylo) ou de souvenirs de problèmes du même type que celui proposé.

Il y a aussi tout un volet de choix stratégiques inconscients déterminés par des « facteurs intrinsèques », c'est-à-dire des caractéristiques de la question ou de la façon dont le problème est posé, avant que toute tentative (ne serait-ce qu'exploratoire) de solution ne commence. Reder a beaucoup étudié le sentiment de familiarité que suscite en nous l'exposition à une question : l'intensité de ce sentiment est très importante pour le choix stratégique. Au niveau intuitif, il s'agit de l'effet du « quiz à la télévision » : tout de suite après avoir lu une question, on se dit « celle-ci je la connais » ou plutôt « non, je ne le sais pas ». On a prouvé que ce sentiment s'active à partir de

⁵¹² L. M. Reder et C. D. Schunn, "Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task," *Cognitive regulation of performance: Interaction of theory and application. Attention and Performance XVII* (1999).

⁵¹³ L. M. Reder, "Selection strategies in question answering," *Cognitive Psychology* 19 (1987).

caractéristiques de la question, non pas de stratégies exploratoires de solution. Par exemple, Reder a fait varier la difficulté des questions et a essayé de manipuler le sentiment de familiarité en stimulant les participants avec des mots qui allaient figurer dans les questions. Comme prévu, la stimulation a eu l'effet de produire un sentiment de familiarité et la conviction qu'on connaissait la réponse, indépendamment du niveau de difficulté (et donc même si les participants se trompaient). Si le sentiment de familiarité avait été produit par une tentative exploratoire de solution, il aurait été sensible à la difficulté des questions⁵¹⁴. Un autre facteur intrinsèque qui influence le traitement d'un problème est la similitude superficielle (au niveau de notation, des tournures mobilisées) avec d'autres problèmes rencontrés par le passé⁵¹⁵.

Globalement, il paraît clair que des fonctions de monitoring et contrôle méta-cognitif plutôt avancées peuvent être automatisées. Le processus d'apprentissage ne fait autre chose qu'alléger progressivement la charge méta-cognitive consciente nécessaire pour accomplir une tâche. Le novice réfléchit énormément et a beaucoup plus de doutes sur l'approche à suivre que l'élève confirmé. Vraisemblablement, pour avancer dans l'apprentissage, il faut un monitoring du taux de succès de plusieurs stratégies et une adaptation conséquente, que cela soit fait consciemment ou non.

La vraie question à poser porte sur les limites dans la quantité des compétences cognitives qu'on peut automatiser de cette façon. Intuitivement, si l'automatisation profonde d'un domaine de compétence entraîne la prise en charge avec succès (par le système cognitif) de la plupart des situations qui peuvent se présenter, il semble qu'un domaine trop vaste, ouvert et dont les limites sont mal définies ne pourra pas être entièrement automatisé. L'automatisation n'est pas une formule magique : il est le résultat d'avoir rencontré plusieurs fois un ensemble extrêmement vaste de circonstances finement déterminées. La quantité absolue de compétences automatisées peut augmenter indéfiniment, mais si l'ensemble de situations qu'il faut savoir gérer avec aisance (pour garder un niveau d'excellence) est suffisamment vaste et se renouvelle rapidement, l'automatisation ne pourra pas tout gérer. Prenons le cas d'un champion de tennis qui veut garder son niveau : la masse de nouveautés techniques et stratégiques introduites par les nouveaux arrivés dans le circuit international est telle que, à chaque moment donné, le champion ne pourra pas avoir tout intériorisé. Il devra aussi réfléchir et délibérer. En plus, dans la performance excellente l'attention consciente et

⁵¹⁴ Ibid.

⁵¹⁵ D. A. Hinsley, J. R. Hayes et H. A. Simon, "From words to equations: Meaning and representation in algebra word problems," *Cognitive processes in comprehension* (1977).

la délibération ne retrouvent pas simplement leurs rôle et forme habituels. On serait tenté de penser que la psychologie d'un expert confronté à un cas extrêmement difficile redevient semblable à celle du novice. Les deux sont contraints à un effort d'attention et de délibération, la seule variable étant le niveau de difficulté nécessaire pour mettre à l'épreuve les compétences acquises. Il n'en est pas ainsi : un changement qualitatif intéressant se produit. Pour comprendre cela, il faut introduire la question de la flexibilité associée à l'excellence.

Flexibilité et excellence

Une marque de la distinction entre maîtrise et compétence est la flexibilité⁵¹⁶. L'automatisation presque complète propre à la maîtrise a un prix à payer. Comme le dit Sternberg : « *l'expertise a aussi bien des coûts que des bénéfices. L'un des coûts est la rigidité croissante : l'expert peut se figer sur un point de vue ou sur une procédure, si bien qu'il devient difficile pour lui de voir les choses différemment* ⁵¹⁷ ». L'expert ne réagit pas uniquement de façon différente et qualitativement supérieure : il est aussi beaucoup plus rapide dans ses choix interprétatifs. Indépendamment de l'efficacité, l'automatisation donne moins de recul par rapport à chaque décision. Ce phénomène est connu comme « effet *Einstellung* » dans la littérature psychologique⁵¹⁸. L'effet *Einstellung* se produit lorsqu'une solution familière (pour l'expert) n'est pas appropriée, à cause de modifications subtiles qui n'attirent pas l'attention. L'expert se trompe : il tombe dans le piège de la familiarité et de l'habitude.

La littérature sur la psychologie de l'automatisation confirme l'intuition et prédit des effets de ce genre. Comme nous l'avons vu, plusieurs théories soulignent la puissance du « vocabulaire perceptif » (*chunks*) de l'expert, ainsi que l'efficacité de la liaison qui est établie entre la situation et les données pertinentes en mémoire de travail à long terme. L'expert saisit correctement la situation et accède rapidement à des données très bien indexées. La qualité de la perception et de l'indexation est proportionnelle à la qualité de la performance. D'autres théories mettent l'accent sur la transformation de connaissance déclarative en connaissance procédurale. Cette conversion serait progressive, avec des items de contenu déclaratif transformés en de simples procédures, qui

⁵¹⁶ Bilalic, McLeod et Gobet, "Inflexibility of experts—Reality or myth? Quantifying the *Einstellung* effect in chess masters."

⁵¹⁷ R. J. Sternberg, "Costs of expertise," in *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*, édition K. A. Ericsson (Hillsdale, NJ: Lawrence Erlbaum Associates, 1996), 347.

⁵¹⁸ A. S. Luchins, "Mechanization in Problem Solving; the Effect of *Einstellung*," (1942). Le mot *Einstellung* se réfère à l'attitude, au « cadre interprétatif » par rapport à un stimulus.

sont ensuite agrégées en procédures plus complexes. Finalement, les procédures qui marchent mieux seraient davantage « renforcées » et deviendraient plus accessibles dans la mémoire. L'automatisation est complète lorsque les situations sont associées à des procédures pertinentes et performantes⁵¹⁹.

Il existe des théories qui définissent un processus automatique comme un processus guidé par l'activation des souvenirs d'expériences passées spécifiques (*past instances*)⁵²⁰. L'action peut être guidée soit par un algorithme général (règles, principes), soit par une comparaison avec des expériences passées sauvegardées dans la mémoire. Pour pouvoir abandonner l'algorithme et passer en mode « analogique », il faut posséder un nombre conséquent d'expériences de résolution du problème. L'automatisation est complète lorsque l'expérience passée la plus pertinente s'active toujours. Dans cette approche, l'expert s'est *littéralement* retrouvé très souvent dans des situations extrêmement semblables à celle à laquelle il est confronté.

Des effets *Einstellung* sont prévus par toutes ces théories. Qu'il s'agisse d'une perception extrêmement puissante et spécialisée (associée à des données bien indexées dans la mémoire de travail), de connaissances procédurales ou d'expériences passées, l'automatisation comporte l'activation de plus en plus rapide de structures cognitives élaborées. Les critères à l'œuvre relèvent de la reconnaissance, de l'analogie ou de la similitude : cela ne fait qu'augmenter le risque de tomber dans le piège de la familiarité. Une preuve de la rigidité associée à l'expertise est la suivante : les experts dans un jeu ont souvent plus de mal que les novices à s'adapter à une variation significative des règles⁵²¹. Par exemple, les serveurs, malgré leur expérience, ont plus de difficulté que les autres à juger l'angle formé par le verre et la surface du liquide, lorsque le verre lui-même est incliné (leur expérience porte sur des verres qui ne sont pas normalement inclinés)⁵²².

Ces données sont indirectes, car elles ne concernent pas des pièges dans l'exercice normal de l'expertise (le bridge modifié n'est pas du bridge). Mais des données directes ne manquent pas : des fautes dues à l'excessive familiarité sont très répandues chez les éditeurs professionnels de textes ; plusieurs accidents graves – notamment dans le nucléaire – sont dus à l'incapacité de repérer des signaux de danger dans une situation qui semble rentrer dans un cadre de figure très familier. Le cas

⁵¹⁹ J. R. Anderson, "Automaticity and the ACT-Super (*) theory," *American Journal of Psychology* 105, no. 2 (1992).

⁵²⁰ G. D. Logan, "Toward an instance theory of automatization," *Psychological Review* 95, no. 4 (1988).

⁵²¹ Pour un exemple avec le bridge, voir P. A. Frensch et R. J. Sternberg, "Skill-related differences in game playing," in *Complex problem solving: Principles and mechanisms*, édition P. A. Frensch et R. J. Sternberg (Hillsdale, NJ: Lawrence Erlbaum Associates, 1991).

⁵²² H. Hecht et D. R. Proffitt, "The price of expertise: Effects of experience on the water-level task," *Psychological Science* 6, no. 2 (1995).

le plus étudié d'effet *Einstellung* est le jeu des échecs. Ici, on a une vision détaillée du phénomène. Si personne n'est à l'abri de ces pièges dans des cas extrêmement difficiles⁵²³, les experts qui se situent jusqu'à trois déviations standards au-dessus de la moyenne dans les classements officiels en sont des victimes très fréquentes : leur performance retombe au niveau d'un joueur moyen. Par contre, les très grands champions de niveau international répondent presque toujours correctement. La performance des joueurs excellents est intéressante parce qu'elle inverse la tendance en place dans le processus d'apprentissage depuis le stade du novice. On allait de l'attention consciente et l'effort du débutant à une approximation de l'automatisme pré-consciente chez l'expert. L'expert travaille « en pilote automatique ». Or, pour éviter des effets *Einstellung*, l'expert excellent est contraint à reprendre contact avec ses propres automatismes et à garder un peu de recul. Il ne s'agit pas nécessairement d'un retour à la délibération explicite, mais sans doute d'une augmentation du contrôle conscient. On retrouve chez l'expert excellent un contrôle fin et un calibrage exquis de ses propres automatismes. Cela génère des réactions créatives et adaptées à des situations nouvelles : « *Les dactylos experts sont capables de s'arrêter deux frappes après qu'ils ont commis une faute, alors que leur performance est censée être automatisée et donc en dehors du contrôle conscient (...)* *Les experts de ping-pong arrivent à ajuster leur frappe 150 ms avant le contact avec la balle (...)* *Il semble que même dans des domaines qui sont plus susceptibles à l'automatisation que les échecs, les experts, au moins les meilleurs entre eux, sont capables d'un contrôle rapproché de leur performance, restant ainsi flexibles*⁵²⁴. »

Les automatismes de l'expert excellent sortent de la boîte noire : il ne s'agit plus d'une deuxième nature dont on n'a pas conscience ou d'un pilote automatique auquel on confie la responsabilité de l'implémentation de nos plans. Ils sont finement intégrés dans la personnalité de l'expert et leur mobilisation est orchestrée de façon consciente. Ils sont aussi constamment retravaillés, car le maintien et la progression dans l'excellence demande un entraînement délibéré continu. Surtout, comme la psychologie de l'expertise le montre de façon claire, un tel niveau d'excellence n'est atteignable que dans un nombre très limité de domaines, et presque toujours dans un seul.

⁵²³ Par exemple, il s'agissait de tester si le joueur s'apercevait qu'une stratégie non optimale dans une position donnée (elle aurait été optimale, mais des détails subtils dans la configuration la rendaient en fait impraticable) *redevient* optimale dans une position très semblable. Voir Bilalic, McLeod et Gobet, "Inflexibility of experts—Reality or myth? Quantifying the *Einstellung* effect in chess masters," 92.

⁵²⁴ *Ibid.*: 96.

L'intégration des vertus et le rôle de la sagesse

Introduction

La vertu est une forme d'excellence pratique en plusieurs domaines. À l'aune des données empiriques analysées jusqu'à présent, il ressort que l'excellence pratique, surtout dans les domaines sensibles pour la morale, est extrêmement difficile à atteindre. À cause de l'état « naturel » de fragmentation du caractère humain et de la tendance à succomber à d'innombrables pressions situationnistes, un comportement moral très cohérent relève de l'excellence, non de la simple compétence. La compétence est un niveau d'expertise qui peut être atteint par tout le monde avec une quantité raisonnable d'effort. Par exemple, une compétence est savoir conduire une voiture, ou jouer « décentement » le piano. L'excellence demande un travail très long et difficile de « pratique délibérée », une forme d'entraînement mentalement très exigeant et soigneusement planifié pour améliorer les points faibles dans sa performance. De plus, l'excellence en plusieurs domaines est très rare. Les transformations cognitives induites par la pratique délibérée sont très profondes et, pour cette raison, l'excellence dans un domaine pourrait même constituer un handicap pour atteindre l'excellence dans d'autres domaines.

Il s'agit d'objections très fortes contre la théorie de la vertu, qui est fondée sur l'idée que plusieurs formes d'excellence pratique interagissent et sont intégrées, notamment dans le domaine de la vie morale.

Or, mettre en avant le rôle intégrateur joué par la notion de sagesse est exactement la stratégie de la théorie de la vertu pour contrer ces objections. La sagesse est une source d'intégration de l'excellence pratique entre vertus différentes: le sage sait comment résoudre les conflits entre le courage, la justice et la modestie. Il est possible que, dans les domaines d'excellence pratique qui sont moralement neutres, on ne dispose pas des avantages offerts par la sagesse. La sagesse ne peut pas aider un physicien excellent à être aussi un écrivain excellent, ni un médecin spécialisé dans la chirurgie nasale à être un médecin généraliste excellent. Voilà donc la réponse de la théorie de la vertu: dans le domaine moral, la sagesse rend possible une intégration qui reste insaisissable ailleurs. Le présent chapitre est consacré à tester le bien-fondé empirique de cette stratégie de réponse. La conclusion sera négative: à l'aune de nos meilleures connaissances, la sagesse ne semble pas en mesure de résoudre le problème de la « fragmentation de l'excellence », qui est aussi grave dans le domaine moral qu'ailleurs. Loin de pouvoir résoudre le problème de la fragmentation,

la sagesse est probablement elle-même fragmentée en plusieurs sous domaines (sagesse dans la vie de famille, dans la gestion des relations professionnelles, etc.). Cela ne laisse qu'une possibilité: un modèle de la vertu en tant que spécialisation morale.

SECTION 1 : brève introduction à la notion de sagesse

La notion de sagesse : questions méthodologiques

Le but du présent chapitre est d'évaluer le rôle que la sagesse pourrait jouer dans le cadre du modèle classique « globaliste » du caractère. Notamment, la question que nous intéresse est plus précise : la sagesse nous offre-t-elle une solution au problème de la cohérence du caractère et de la conduite moraux (qui nous a occupé au fil du présent travail) ? Est-ce que la solution au problème de la fragmentation de l'expertise morale réside dans la sagesse ? La sagesse est-elle un facteur d'unification, où il faudra plutôt reconnaître que la sagesse est elle-même une expertise aussi fragmentée (dans plusieurs domaines : sagesse dans la vie publique, dans la vie de famille, dans la recherche intellectuelle) que les autres formes d'expertise ?

Nous visons une question spécifique, non pas l'élaboration d'une théorie générale de la sagesse. Néanmoins, il nous faut commencer par une courte présentation générale de la notion. Elle occupe une place trop importante dans plusieurs traditions philosophiques (occidentale et orientale) et dans l'histoire de l'humanité (ce n'est pas une exagération !), pour qu'on puisse la mobiliser sans une présentation et un état des lieux.

L'histoire de la notion de sagesse est aussi étendue dans le temps et dans l'espace que celles de toute tradition philosophique, religieuse et littéraire qui soit parvenue jusqu'à nous⁵²⁵. La notion de sagesse est extrêmement riche et a plusieurs facettes : il est difficile de définir les contours de son histoire de façon rigoureuse, tout en évitant l'éclatement total du débat. C'est un problème qui concerne les notions de vertu en général (et encore plus généralement, les notions centrales du sens commun) : par exemple, somme nous certain que le courage dans la bataille, le courage dans la vie politique, le courage de prendre la parole dans une réunion d'entreprise (etc.) aient suffisamment de

⁵²⁵ La littérature sur la sagesse - composée essentiellement de recueils de maximes, histoires exemplaires et proverbes, ainsi que de réflexions inspirées par ce matériel - est très ancienne. Dans la tradition occidentale, les références évidentes sont l'Ancien Testament (Proverbes, Job, Livre de la Sagesse) et les classiques grecs et romains. Mais la littérature sur la sagesse inclut des textes venant d'Egypte, Perse (Zoroastrianisme), Mésopotamie, Inde (Bouddhisme) et Chine (Confucianisme et Taoïsme). Pour une histoire détaillée de la littérature ancienne sur la sagesse, voir Baltes, "Wisdom as Orchestration of Mind and Virtue".

facteurs en commun pour qu'on puisse envisager une discussion générale sur le courage, et une théorie du courage ? Ne vaut-il mieux admettre qu'il y a (et il y a eu, au fil de l'histoire, dans des contextes théoriques différents) plusieurs débats distincts et vaguement homologues, si bien qu'aujourd'hui – si on veut être rigoureux – on doit *choisir* la notion de courage qui nous intéresse, et abandonner la question générale ?

Il est vraiment difficile d'éviter l'éclatement du débat sur la vertu : dans le présent travail (notamment dans l'introduction), nous avons défendu une unité minimale mais substantielle de la théorie de la vertu. Depuis deux milliers d'années, les philosophes débattent de la possibilité de transformer de façon ambitieuse le caractère humain, en sorte qu'il manifeste de façon pleine les qualités exprimées par le vocabulaire normatif des traits. En bref, la théorie de la vertu est une théorie normative ambitieuse du caractère, basée sur les termes de trait. Cette définition paraît extrêmement minimale mais elle n'est pas vide. Les termes de trait sont une classe de termes évaluatifs qui existe dans toutes les cultures ; ils ont une mode de fonctionnement caractéristique : chacun d'entre eux a un noyau sémantique minimal qui permet de les distinguer d'autres concepts du sens commun (voir le chapitre sur la théorie des traits du sens commun). Malgré la grande quantité de choix théoriques qui restent ouverts, les théories de la vertu qui se sont succédées au fil de deux mille ans d'histoire *sont* en compétition entre elles. Par exemple, les théoriciens de la vertu que nous avons qualifié de « déflationnistes »⁵²⁶ nous doivent expliquer pourquoi l'idéal traditionnel d'un caractère exceptionnellement cohérent et harmonieux est à abandonner. Ce ne serait pas correct d'interpréter la décence morale (un niveau moyen et acceptable de moralité) et l'excellence morale comme deux sujets d'étude indépendants. Les théories de la vertu comme « décence » et les théories de la vertu comme « excellence » sont en compétition, au niveau de la façon dont elles exploitent le potentiel normatif des notions de trait.

Ce type d'unité minimale mais substantielle représente la bonne approche pour reconstituer le débat sur la sagesse.

Il est impossible de formuler le problème de la sagesse d'un point de vue théoriquement neutre : la notion de sagesse est extrêmement riche et présente plusieurs facettes. Toute discussion rigoureuse doit faire des choix de définition qui donnent une forme précise au problème : il se peut qu'un penseur se concentre sur une dimension qu'un autre penseur ignore ou exclut, directement à partir de la façon dont ce deuxième formule la question. Par exemple, partir d'une distinction entre sagesse pratique (savoir faire et jugement « sage » dans la solution de problèmes pratiques de la

⁵²⁶Voir le chapitre sur le modèle néo-aristotélicien.

vie : relations, équilibre entre les différents domaines de la vie) et sagesse théorique ou épistémique (capacité de discernement de jugement « sage » appliqués à la solution de problèmes théoriques) pourrait sembler raisonnable : après tout, l'enquête rationnelle et la vie pratique constituent deux domaines bien distincts de l'existence humaine, dans lesquelles le jugement sage peut s'exercer. Cette distinction est un point de départ très fréquent. Mais bien de théoriciens ont défendu l'unité de ces deux formes de sagesse, mettant en avant une conception rigoureusement unitaire de la sagesse⁵²⁷.

Nous voudrions renverser cette forme de paralysie de l'analyse et de pessimisme (caractéristique de toute notion importante tirée du sens commun), en faveur d'un optimisme inspiré par une approche taxinomique et historique : malgré la variation spatiale et temporelle, la notion de sagesse présente un noyau sémantique minimal remarquablement informatif et universel. Étant donné que le débat sur la sagesse est transhistorique (plusieurs millénaires), trans-culturel et trans-disciplinaire (philosophie, religion, culture populaire), c'est étonnant que, à chaque fois, il y ait été question de notions qui se ressemblaient à ce point ! Nous allons procéder à articuler ce noyau minimal, puis à exposer certaines variantes plus locales de la notion, qui correspondent à des choix théoriques importants. Ensuite, nous allons nous intéresser à la notion de sagesse dans le sens commun.

La double nature de la sagesse : le « jugement sage » et la « personnalité du sage »

En ce qui concerne le noyau universel minimal du concept de sagesse, il est essentiel de commencer par une distinction entre deux sens de sagesse : la sagesse en tant que capacité de « jugement sage » (dans cette notion, il faut inclure toutes les formes de connaissance tacite ou implicite qui contribuent à cette capacité de jugement) et la sagesse en tant que personnalité harmonieuse, propre aux individus qui ont complètement intériorisé et qui mettent constamment en pratique les enseignements issus de leur capacité de jugement sage. Autrement dit, la sagesse est une forme « d'expertise dans le jugement » qui transforme complètement la personnalité de qui l'applique. La sagesse est une forme de connaissance supérieure (dans un sens très large de connaissance) qui amène le caractère de l'être humain à un niveau d'existence supérieure. Si on devait résumer dans une phrase le sujet de plusieurs millénaires de débat, nous croyons que ce serait une formulation adéquate.

⁵²⁷ L. T. Zagzebski, *Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge University Press, 1996). Voir ci-dessous.

En soi, la distinction entre « jugement sage » et personnalité du sage n'est pas difficile à comprendre. Dans le cadre de toute forme d'expertise, on peut distinguer entre la capacité de jugement et les compétences nécessaires pour une action efficace. Un juge expert, que ce soit dans le domaine du tennis, de la poésie ou de la musique jazz, n'est pas forcément en mesure de faire ce dont il est capable de juger la qualité. La distinction, appliquée à la sagesse, est exactement la même. Par contre, il y a une objection importante : les deux notions sont probablement liées. Est-ce que la distinction tient vraiment la route ? Les deux notions ne sont-elles indissociables ? Ne s'agit-il pas d'une seule compétence ?

Il y a deux liens possibles entre le jugement et la mise en pratique : « jugement entraîne mise en pratique » et « mise en pratique entraîne jugement ». D'après le premier lien, un bon juge doit avoir pratiqué. Sans doute, il faut un certain niveau de compétence opérationnelle pour avoir un niveau de jugement intéressant. Il est difficile d'imaginer qu'un très bon entraîneur de tennis puisse ne pas avoir été un joueur d'un niveau significatif, même si pas un champion. D'après le deuxième lien, ce n'est pas possible d'avoir un certain niveau de performance (en pratique), si l'on n'a pas le niveau de jugement « correspondant ». Ce n'est pas possible d'avoir un niveau « confirmé » en improvisation théâtrale sans la capacité de discriminer entre la prestation d'un débutant et une prestation d'un niveau plus élevé.

Nous allons revenir ci-dessous sur le lien effectif entre jugement et performance. La question est cruciale car, nous le verrons, la psychologie de la vertu classique est basée sur l'hypothèse d'une corrélation parfaite entre les deux facteurs : il n'y a pas d'action vertueuse sans jugement vertueux, ni de jugement réellement vertueux chez quelqu'un qui ne le mettrait pas en pratique. Pour l'instant, il nous suffit de répondre que, alors même que le jugement et l'action étaient parfaitement corrélés, on les a toujours distingué au moins d'un point de vue conceptuel. Dans toutes les traditions philosophiques et non, la sagesse a été conçue comme une forme de jugement *suis generis*, capable de transformer le caractère. Le processus de transformation est très long, c'est le travail d'une vie. Par exemple, vers la fin de sa vie, Confucius semble avoir dit : « À 15 ans j'ai commencé à apprendre, à 30 j'étais confirmé, à 40 ans je n'avais plus de doutes, à 50 ans je connaissais la volonté du Ciel, à 60 ans j'étais prêt à l'écouter, à 70 ans je pouvais suivre mes désirs sans transgresser ce qui est juste⁵²⁸ ».

La dimension de savoir (au sens très large) peut être partiellement « isolée » : la sagesse peut être « stockée » dans des livres et dans des recueils de proverbes. Même si la sagesse d'un livre sacré (par exemple) ne devient vivante que chez quelqu'un qui sait l'interpréter proprement, elle a au moins

⁵²⁸ Cité dans Baltes, "Wisdom as Orchestration of Mind and Virtue", 50.

une vie partiellement indépendante, ne serait-ce que parce que, au fil de l'histoire, on a souvent cru qu'elle vient des dieux, d'une révélation, d'autres maîtres ou d'une tradition. En tout cas, le sage doit acquérir un savoir qui lui est transmis. D'après Confucius : « *Pour apprendre à être soi-même, il faut avoir accès à un vaste ensemble de connaissances portant sur ce qui a été dit et fait dans le monde ; il faut l'analyser de façon critique, y réfléchir longtemps, le passer au crible avec un esprit lucide, et le mettre en pratique*⁵²⁹ »

Le thème de la sagesse a été presque universellement introduit suivant le schéma d'une forme de jugement (qui mobilise un savoir spécifique et *sui generis*) qui, venant à contact d'un individu humain, est en mesure de transformer et réorganiser complètement son caractère, jusqu'à un niveau d'existence supérieure.

Le noyau sémantique minimal de la sagesse en tant que jugement sage

Quelle est la nature de la sagesse en tant que jugement sage⁵³⁰ ? Parmi les thèmes (presque) universels, on retrouve sans doute les suivants⁵³¹ :

1. la sagesse s'occupe de questions importantes et difficiles portant sur la conduite humaine et la signification globale de la vie
2. la sagesse inclut des (méta-)connaissances sur les limites de la connaissance humaine et sur la gestion de situations incertaines.
3. la sagesse représente un niveau supérieur de connaissance, de jugement et de capacité à donner des conseils.
4. la sagesse constitue une forme de connaissance extrêmement profonde, équilibrée, holiste et pertinente dans un ensemble très large de situations.
5. la sagesse contribue à orchestrer et intégrer de façon harmonieuse la personnalité (en ce sens, elle joue un rôle métacognitif d'arbitrage entre d'autres compétences cognitives et pratiques)
6. la sagesse est un savoir au service du bien-être individuel et collectif
7. la sagesse, difficile à définir et à acquérir, est facilement reconnaissable dans la vie de tous les jours.

⁵²⁹ Cité dans Ibid.

⁵³⁰ Dans cette section, par sagesse nous allons entendre toujours jugement sage. Lorsque nous disons que la sagesse est un savoir, nous nous référons au corpus d'informations et de connaissances tacites qui sont mobilisées au service du jugement sage.

⁵³¹ Notre source principale est l'étude sur l'histoire de la notion de sagesse faite par Baltes, voir Baltes, "Wisdom as Orchestration of Mind and Virtue", 17.

Une notion de sagesse ayant au moins ces caractéristiques se trouve au centre de presque toute grande tradition philosophique, littéraire ou artistique du monde⁵³². Il est difficile de juger si le niveau de diffusion d'une notion est surprenant ou non, car il est difficile d'avoir de termes de comparaisons pertinents, mais retrouver toutes ces propriétés dans autant de contextes différents est sans doute remarquable.

Plusieurs aspects de la notion de sagesse méritent d'être soulignés. D'abord, la sagesse est un savoir qui a une dimension « cosmique » et qui concerne les aspects les plus fondamentaux de la place que l'homme occupe dans la réalité (1). Il ne s'agit pas pourtant d'une métaphysique ou d'une cosmologie qui décriraient la structure de l'univers dans tous ses détails, y compris l'homme et son existence. Il y a deux différences entre la sagesse et ce projet encyclopédique de description de la réalité. Premièrement, la notion pertinente d'univers et de cosmos n'est pas la même : la sagesse s'intéresse à l'univers en tant que champ d'action possible de l'homme, véritable « scénario de jeu », déterminé de la façon la plus large possible. Il y a d'innombrables aspects de la réalité qui ne relèvent pas du niveau d'analyse anthropocentrique de la sagesse : par exemple, la composition chimique des roches ou la fusion dans le noyau des étoiles⁵³³. Deuxièmement, la sagesse n'est pas avant tout un savoir positif et affirmatif. Ce côté existe, bien sûr : la littérature sur la sagesse abonde en prescriptions pratiques spécifiques à la culture ou plus universelles, telles que « Ce qu'on ne désire pas pour soi, ne pas le faire à autrui ». Mais la partie affirmative et « substantielle » de la sagesse n'est pas ce qui fait sa spécificité : la sagesse concerne les questions ultimes de l'existence humaine tout en restant bien sensible aux limites de la « situation épistémique » de l'homme (2).

Il serait plus exact de dire que la sagesse est un savoir « adapté » aux aspects les plus profonds de la condition humaine, vu qu'il ne s'agit pas d'une anthropologie philosophique positive (un savoir substantiel « sur » la condition humaine) ni d'un « catéchisme moral » ou prudentiel⁵³⁴. Il est question de ce qu'on ne peut pas savoir, de ce qu'on peut difficilement savoir, de la gestion du risque et de l'incertitude. Par conséquent, la sagesse est un savoir fluide, heuristique, qui prépare à l'action plutôt que donner de réponses théoriques précises.

⁵³² Ibid.

⁵³³ C'est difficile de trouver des exemples clairs d'aspects de la réalité qui ne sont pas pertinentes du point de vue de la sagesse : après tout, la fusion dans les étoiles pourrait être importante pour comprendre les origines du cosmos et pour délimiter le champs d'action possible de l'homme. En tout cas, la sagesse ne concerne pas les détails « techniques » du monde, mais la monde en tant que « scénario de jeu » pour l'homme.

⁵³⁴ Au contraire, une critique fréquente aux catéchismes trop substantiels est la perte de toute dimension de sagesse.

Cela ne relève pas pour autant du simple questionnement, d'une mise en perspective critique, voire sceptique. Il ne s'agit pas d'alimenter un enrichissement réflexif de notre compréhension du monde, sans garantie sur l'issue du processus. Qui profite au plein de la sagesse est extrêmement efficace. Le sage est recherché pour son jugement et ses conseils pratiques (3). Il est vrai que la sagesse est souvent considérée comme insaisissable, mystérieuse, voire mystique. Le jugement et le conseil du sage nous surprend et nous dépasse dans sa profondeur. Mais cette dimension mystérieuse de la sagesse n'est pas aussi universelle qu'on pourrait le croire : elle évoque l'image d'un sage zen, mais, dans d'autres traditions un exemple typique de sage a été, par exemple, un général ou un grand homme d'état, des figures beaucoup moins mystiques.

« C'est pourquoi nous estimons que Périclès et les gens comme lui sont des hommes prudents en ce qu'ils possèdent la capacité d'apercevoir ce qui est bon pour eux-mêmes et ce qui est bon pour l'homme en général, et tels sont aussi, pensons-nous, les personnes qui s'entendent à l'administration d'une maison ou d'une cité.⁵³⁵ »

Surtout, le caractère mystique risque de mettre au deuxième plan l'efficacité, qui est un trait beaucoup plus important de la sagesse. Le sage n'est pas un gourou (qui est tel uniquement parce qu'il est *perçu* comme étant profond par un certain nombre de personnes), ses conseils sont avant tout efficaces, d'une manière (au moins potentiellement) testable. Le conseil d'un sage permet de résoudre une crise diplomatique, un problème relationnel ; le sage arrive à convaincre un roi de changer un politique néfaste⁵³⁶ ; si son succès n'est pas évident de son vivant, ce sera l'histoire à lui donner raison. Certes, la façon dont il arrive à faire cela semble montrer des capacités supérieures et engendre une réaction de mystère chez qui le voit en action, mais l'action du sage est avant tout définie par son efficacité extrême et concrète.

La sagesse est fluide et heuristique : sa mobilisation rend le sage très opérationnel. En tant que savoir « adapté » et sensible aux aspects les plus profonds de la condition humaine, la sagesse est par nature *holiste* (4). Le sens concerné de holisme est le suivant : on peut certainement être plus au moins sage et peut-être plus sage dans un domaine que dans l'autre, mais la sagesse n'est jamais strictement limitée à un domaine d'action (au moins, en ce qui concerne la notion traditionnelle de sagesse). Un choix très intelligent et même astucieux du point de vue professionnel mais qui sacrifierait complètement la vie privée n'est pas sage.

⁵³⁵ 1140b Aristote, *Ethique à Nicomaque* 306. Sagesse pratique (*phronesis*) est traduit par « prudence ».

⁵³⁶ Dans la conception chinoise de la sagesse, l'efficacité est beaucoup plus importante que le caractère mystique et contre intuitif de l'action du sage. Les sages jouaient un rôle très important de médiateurs et de conseillers politiques. Voir Kathleen Marie Higgins, "Negative virtues: Zhuangzi's Wuwei," in *Virtue Ethics Old and New*, édition Stephen M. Gardiner (Cornell University Press, 2005).

La sagesse n'est jamais « aveugle » à aucun secteur important de la vie : la sagesse fixe des priorités, met en perspective et aide exactement dans le genre de choix qui opposent vie professionnelle, familiale ou privée. Les solutions proposées ne visent jamais le simple intérêt individuel, pas même dans une conception sophistiquée et élargie où le bien-être des proches serait comptabilisé dans l'épanouissement d'une vie individuelle.

Même si les textes de la tradition n'utilisent pas ce terme psychologique contemporain, la sagesse est très fortement *méta-cognitive*. Elle l'est au moins en deux sens : elle joue un rôle d'arbitrage entre émotions, exigences, habitudes et processus psychologiques différents. Une décision sage (p.ex. accepter un poste de responsabilité au service de sa communauté) trouve un équilibre entre peurs, espoirs, intérêt personnel et devoirs d'un individu. La sagesse est sensible à l'état émotionnel et cognitif de l'individu, d'une façon assez fine. À côté de cette gestion méta-cognitive de l'état présent de l'individu en vue d'une évolution à court ou moyen terme, les décisions sages sont également adaptées aux limites de la cognition humaine à très long terme (point 2 de la liste, et deuxième sens de « méta-cognitif ») : comme nous l'avons déjà souligné, la sagesse est sensible au « scénario de jeu » de l'homme à une échelle cosmique. Pour ainsi dire, les arbitrages et les équilibres émotionnels et cognitifs favorisés par la sagesse sont « soutenables » et efficaces à très long terme.

La sagesse vise toujours l'équilibre entre l'intérêt individuel et collectif, compris en un sens beaucoup plus large (6) : le sage est sensible aux conséquences (symboliques ou factuelles) de son action à l'échelle de la société, d'une nation, voire du genre humain.

Ces propriétés définissent le noyau universel minimal de la sagesse en tant que jugement sage, mais il ne faut pas oublier que cette notion n'appartient pas uniquement à des traditions philosophiques ou religieuses élaborées : il s'agit d'un concept du sens commun (7). La sagesse est un concept opérationnel, facilement reconnaissable dans ses manifestations pratiques (l'identification des sages ou d'actions sages) et plus au moins mobilisé de manière active, selon la culture et la période historique. C'est là une autre preuve du fait que le débat sur la vertu appartient au sens commun⁵³⁷.

La sagesse en tant que personnalité vertueuse du sage

Or, l'aspect le plus controversé et important de cette notion clé du sens commun et de l'histoire de la civilisation humaine est sa fonction dans la construction et dans la stabilisation d'une personnalité

⁵³⁷ Une section entière de ce chapitre sera consacrée à l'analyse de la notion de sagesse du sens commun, et de ses liens avec la notion de sagesse mobilisée par les traditions philosophiques et religieuses.

épanouie : en d'autres termes, le rôle de la sagesse en tant que jugement dans l'intégration et l'orchestration des vertus. Le type de caractère que le jugement sage contribue à développer, chez quelqu'un qui met en pratique et cultive ses enseignements, n'est autre chose que la vertu pleine. La personnalité du sage est tout simplement la personnalité de l'individu pleinement vertueux. L'individu vertueux est quelqu'un qui met en pratique les enseignements du jugement sage et a enduré le processus de transformation du caractère que cela entraîne. La sagesse joue un rôle clé dans la théorie de la vertu : nous consacrerons plusieurs pages à l'histoire de ce rôle ci-dessous. Pour l'instant, étant donné que nous nous intéressons au noyau sémantique minimal de la notion de sagesse, revenons au problème du lien entre jugement et mise en pratique. Est-il possible que la capacité à juger de façon sage et la mise en pratique soient dissociées ?

Le jugement sage est beaucoup plus qu'un savoir froid, déclaratif et abstrait, qui pourrait donc être bien évidemment dissocié de sa mise en pratique : la possession de la sagesse se manifeste au moins (par définition) dans une capacité de jugement efficace et dans la mobilisation d'heuristiques fluides et appropriées. La possibilité de réduire la sagesse à la simple connaissance de quelques proverbes, maximes et citations et à la capacité à les ressortir au moment opportun est exclue. Ce n'est pas de cela que l'on est en train de discuter : l'art d'une conversation efficace n'est pas une manifestation de sagesse. La sagesse est un corpus de savoir dont la simple possession entraîne une série de changements cognitifs et émotionnels profonds. Autrement dit, le développement du jugement sage affecte nécessairement le caractère d'un individu, dans une certaine mesure.

Une connaissance fluide, assortie du déploiement d'heuristiques sophistiquées et de réactions émotionnelles adéquates, semble un candidat idéal pour jouer un rôle dans l'intégration de la vertu. Néanmoins, le développement d'une personnalité vertueuse et harmonieuse est un objectif extrêmement ambitieux. Cependant, un savoir riche et une capacité de jugement raffinée n'assurent pas la cohérence de la conduite. Toute progression au niveau de sagesse (savoir) va-t-elle entraîner une harmonisation et intégration du caractère ? C'est le pari, comme nous le verrons ci-dessus, de la théorie de la vertu classique.

Il ne s'agit pas de la seule tradition qui a lié très étroitement les deux notions de sagesse. Dans les traditions orientales tout autant que chez Aristote, l'indécision, l'inaptitude ou tout simplement une conduite non-excellente dans un contexte d'action sont incompatibles avec toute attribution de sagesse, aussi bien au sens de savoir que de vertu accomplie. Il n'y a pas de savoir sans cohérence dans l'action, ni cohérence dans l'action sans savoir. Se référant à la vertu de la bienveillance (*Ren*), Confucius écrit :

« Si un gentilhomme oublie la bienveillance, comment pourrait-il mériter ce titre ? Un gentilhomme n'agit jamais à l'encontre de la bienveillance, ne serait-ce que pendant le temps nécessaire pour terminer un repas. Même dans une période d'urgence et de souffrance, son action sera en accord avec la bienveillance⁵³⁸ .»

Au premier abord, cela paraît intuitif, car il est difficile d'imaginer un sage incohérent, chez qui l'action ne suivrait pas le jugement. En réalité, la thèse est très radicale. Pour s'en apercevoir, il ne faut pas se concentrer sur le cas de quelqu'un capable de donner de bons conseils mais qui les ignore complètement dans la pratique, jusqu'au point où sa conduite morale serait inférieure ou égale à la moyenne de la population. Il est clair qu'il ne s'agit pas d'un candidat probable à la condition de sage. La question est de savoir si l'excellence dans la compréhension et le jugement est suffisante pour assurer *l'excellence* et la cohérence dans l'action.

En ce qui concerne l'identification de la sagesse en tant que savoir et de la sagesse en tant qu'excellence pratique, le style de contre-exemples pertinents à prendre en considération est plutôt le suivant : un entraîneur excellent, mais qui n'a pas fait les meilleurs choix en tant que joueur et n'a pas toutes les qualités requises pour jouer au niveau international dans sa discipline ; un stratège et conseiller militaire très expérimenté, mais qui ne serait pas à la hauteur d'un poste de responsabilité directe et ne voudrait pas être général. Or, un exemple très typique d'individu sage chez Aristote est un général ou un grand homme d'État. La possibilité de ces contre-exemples montre que l'identification traditionnelle de savoir, personnalité et action « sages » est en réalité un choix théorique radical et ambitieux. Cherchons à suivre l'histoire de cette stratégie théorique, dans le cadre de la psychologie classique des vertus.

Le rôle unificateur de la sagesse dans la psychologie classique des vertus

La sagesse joue un rôle crucial dans la coordination et l'unification des vertus. Pour comprendre ce rôle, il faut commencer par la doctrine de l'unité de la vertu, qui occupe une place centrale dans l'histoire de la théorie de la vertu, à partir de la philosophie grecque ancienne.

La réflexion théorique de l'époque s'est développée dans un milieu culturel où la notion de vertu était beaucoup plus populaire qu'aujourd'hui. Néanmoins, cette notion était sans doute plurielle : il était courant de discuter du rôle du courage, de la justice et de la tempérance dans l'éducation des enfants ou de se demander quelles sont les vertus essentielles pour un bon citoyen. Or, le débat philosophique ne s'intéresse pas *aux* vertus, mais à *la* vertu. Étant donné qu'aujourd'hui la

⁵³⁸ Cité dans Hutton, "Character, Situationism, and Early Confucian Thought," 41.

conception plurielle a regagné le devant de la scène, cette importance accordée au thème de l'unicité par les philosophes paraît quelque peu mystérieuse.

Alors que la théorie de la vertu se proposait de critiquer le sens commun de l'époque, l'idée de l'unicité de la vertu n'était pas motivée par l'intention de mettre en évidence les défauts de la conception pluraliste. L'approche unitaire n'était pas présentée comme la seule façon d'échapper à l'incohérence d'analyses indépendantes de la justice, du courage ou de la tempérance. Ce qui motive est en réalité la question fondatrice de l'éthique ancienne : « Quelle est la vie bonne, excellente, la meilleure qui puisse être vécue ? ». L'hypothèse de travail la plus naturelle pour élucider l'idéal de la vie bonne est qu'il existe une condition unitaire qui en serait responsable, du problème de *la* vertu, et non pas *des* vertus. La vertu est une condition d'excellence qui explique pourquoi la vie de certaines personnes est (ou se rapproche de) la meilleure qui puisse être vécue⁵³⁹. En ce sens, la vertu (au sens unitaire) est introduite comme un « terme de succès ». On peut vouloir découvrir ce que les champions de décathlon ont en commun (qui expliquerait leur performance) ; de la même manière, les philosophes anciens se posaient la question de savoir ce que les vies les plus réussies moralement (un grand général, un grand philosophe, un grand statisticien) ont en commun. Étant donné cette approche théorique, il apparaît que la philosophie s'est intéressée à l'intégration des vertus dès le départ, par *définition* du problème. La question de l'interaction entre plusieurs composants de *la* vertu constitue le point de départ : il ne s'agit pas d'un problème supplémentaire qu'on aurait le droit de poser seulement après avoir étudié séparément la nature de la justice ou du courage.

Dans ce contexte, la liste des vertus du sens commun n'est pas évacuée. L'ambition des théoriciens de la vertu est toujours d'expliquer la nature de l'excellence de ce que nous trouvons courageux, juste ou magnanime. Néanmoins, la tension apparente entre l'unicité de la vertu et la pluralité des termes qui désignent des traits vertueux doit être résolue. Les théoriciens de la vertu anciens ont mis en avant l'hypothèse suivante : tous les termes de vertu se réfèrent à une réalité unique et non structurée.

Il existe deux variantes de cette idée : l'une, extrêmement radicale, soutient la synonymie des termes de vertu. Non seulement la vertu serait unique, mais « courage », « justice » ou « sagesse » seraient de véritables synonymes qui se réfèrent à la même condition psychologique sous-jacente⁵⁴⁰. L'autre, attribuée à Socrate (à partir du Protagoras) et aux Stoïciens Zenon, Ariston et Chrysippe

⁵³⁹ Sur cette reconstitution du débat ancien, voir J. M. Cooper, "The Unity of Virtue," *Social Philosophy and Policy* 15 (1998).

⁵⁴⁰ Position attribuée au stoïcien Ménédème, voir Ibid.

(qui auraient développé la conception présentée dans le Protagoras), insiste sur le fait que la vertu est unique et non structurée, mais admet de variations dans la manifestation. Chez tous ces auteurs, le monisme (non-structuration) de la vertu est assuré par son identification avec la *connaissance* : la vertu n'est autre chose que la connaissance nécessaire pour vivre une vie excellente. On qualifie souvent cette forme de connaissance de sagesse.

Ce savoir est parfaitement unitaire, harmonieux, « holiste ». Pour peu qu'il soit mobilisé, il l'est dans sa totalité, dans toute situation pertinente. Bien évidemment, le contexte et la finalité de l'application peuvent changer, suivant qu'on applique cette connaissance sur le champ de bataille ou dans le commerce. En ce sens, la métaphore qui est évoquée est celle de l'or : sa nature est rigoureusement la même, mais on peut isoler des parties différentes dans une pièce de monnaie. Autre (et meilleure) métaphore : la vue. L'action et le fonctionnement de la vue sont toujours les mêmes, mais on peut observer des scènes différentes. Dans ce cadre théorique, la variation admise est uniquement « relationnelle » et concerne l'application : cela ne touche pas à la nature homogène de la vertu.

Chrysippe abandonne une conception stricte du manque de structuration interne de la vertu : la connaissance du vertueux peut être classée par domaines, qui relèvent des différentes vertus. Il y a du savoir qui a affaire au contrôle des passions, à la justice, etc. Tous ces types de savoir doivent être mobilisés pour chaque action, mais la contribution n'est pas la même : par exemple, si le contrôle des passions est toujours nécessaire, dans un tribunal on fera un usage massif de la connaissance qui relève de la justice. Au contraire, dans la bataille c'est le courage qui l'emporte, même s'il faut aussi être juste. La connaissance qui constitue la vertu reste très holiste chez Chrysippe; elle peut être mobilisée dans des configurations différentes, mais toutes les vertus s'activent pour chaque action vertueuse.

Aristote, dans les interprétations qui paraissent le plus défendables⁵⁴¹, soutient la version la moins radicale de l'unité des vertus. Il existerait plusieurs formes de savoir pratique clairement distinctes. La connaissance nécessaire pour être courageux dans le camp de bataille ou dans la vie politique n'est pas la même connaissance qu'il faut posséder pour être généreux ou juste⁵⁴². Dans des

⁵⁴¹ Par exemple, voir C. Natali, *L'action efficace: études sur la philosophie de l'action d'Aristote* (Peeters, 2004), C. Natali, *La saggezza di Aristotele* (Bibliopolis, 1989).

⁵⁴² Chez Aristote, les vertus correspondent à des macro-domaines de la vie, pas à des domaines spécifiques tels qu'une profession, un rôle social (père, fils, maître). Par exemple, le courage est activé dans tous les domaines où il est question de faire face au risque et au danger (la vie militaire, la vie politique, etc.). Sur cet aspect de l'interprétation d'Aristote, voir Nussbaum, "Non-Relative Virtues: An Aristotelian Approach."

situations non ambiguës, ces compétences pragmatiques différentes pourraient effectivement fonctionner de manière indépendante et autonome. Néanmoins, il y a un problème d'arbitrage. En vue de l'efficacité et d'une conduite moralement irréprochable, il faut pouvoir arbitrer lors d'un conflit parmi les prescriptions de plusieurs corpus de savoir pragmatique. Le corpus de connaissance pragmatique qui gère la conduite courageuse d'un soldat ne contient pas suffisamment d'information sur la justice pour pouvoir décider, dans un cas difficile, si une stratégie d'attaque est appropriée ou injuste et peu honorable, étant donné les normes et les coutumes.

Le savoir de la sagesse contient les connaissances nécessaires pour arbitrer entre différentes vertus dans les cas difficiles et il est nécessaire pour la pleine possession des vertus. On ne pourrait pas considérer comme vertueuse la simple capacité à gérer les cas les plus faciles que l'on rencontre dans les principaux domaines d'expertise pratique de la vie. L'arbitrage est très souvent nécessaire, et la sagesse est indispensable pour la vertu. Mais si la sagesse est un savoir pragmatique de plus haut niveau, qui permet de gérer plusieurs corpus de connaissances pratiques, elle ne peut s'exercer que si l'on possède ces savoirs de base. En ce sens, la vertu est indispensable pour la sagesse: faute d'une certaine quantité de connaissances pratiques pertinentes pour l'exercice de l'honnêteté, du courage, de la magnanimité, de la justice et des autres vertus, la sagesse ne pourrait pas opérer, elle serait privée de sa matière première, de ses inputs.

« On voit ainsi clairement, d'après ce que nous venons de dire, qu'il n'est pas possible d'être homme de bien au sens strict, sans prudence, ni prudent sans la vertu morale. Mais en outre on pourrait de cette façon réfuter l'argument dialectique qui tendrait à établir que les vertus existent séparément les unes des autres, sous prétexte que le même homme n'est pas naturellement le plus apte à les pratiquer toutes, de sorte qu'il aura déjà acquis l'une et n'aura pas encore acquis l'autre. Cela est assurément possible pour ce qui concerne les vertus naturelles ; par contre, en ce qui regarde celles auxquelles nous devons le nom d'homme de bien proprement dit, c'est une chose impossible, car en même temps que la prudence, qui est une seule vertu, toutes les autres seront données.⁵⁴³ »

Aristote défend une forme de holisme psychologiquement plausible, qui reconnaît la spécialisation des connaissances pratiques tout en soutenant la nécessité d'une intégration de plus haut niveau. Quelle est la nature psychologique d'un arbitrage entre vertus ? Chaque vertu étant caractérisée par une façon d'appréhender des situations externes, par des réactions émotionnelles, par des habitudes,

⁵⁴³ 1144B 30–1145a Aristote, *Ethique à Nicomaque* 335-336. Les vertus naturelles sont des dispositions à manifester et acquérir des vertus (par exemple, une certaine attitude naturelle au courage). Sans éducation et entraînement rigoureux, elles ne sont pas suffisantes pour développer la forme « pleine » des vertus correspondantes (le courage naturel n'est pas suffisant pour être vraiment courageux).

des heuristiques et des styles cognitifs, l'arbitrage entre vertu peut avoir lieu à chacun de ces niveaux. Le courage peut me pousser à attaquer un ennemi, alors que la justice favorise une réinterprétation de la même situation selon laquelle je devrai l'épargner (il a été correct et loyal dans le combat) : il est nécessaire un arbitrage, qui peut intervenir déjà au niveau de la perception de la situation. Grâce à la sagesse, le courageux n'aura pas la tentation d'attaquer un ennemi, si cela était injuste. L'arbitrage entre les vertus opéré par la sagesse peut se manifester également plus en aval : par exemple, entre des styles cognitifs ou des habitudes associées à des vertus différentes.

Que la forme d'intégration soit radicale ou modérée, traditionnellement, l'argumentation des théoriciens de la vertu s'appuie sur le rôle d'un savoir fortement holiste dans le bon fonctionnement du caractère vertueux : cette forme de savoir est la sagesse. Le sens commun pense que l'on pourrait être courageux sans être honnête ou juste, généreux sans être courageux ou loyal, etc. Les philosophes objectent que la vertu sans sagesse n'est pas possible et que le développement de la sagesse entraîne l'intégration des vertus. Après avoir terminé la présente introduction de la notion de sagesse, il nous faudra examiner le bien fondé de la conception de la sagesse comme facteur d'unification entre les vertus, à la lumière des données empiriques dont nous disposons. Si la conception traditionnelle de la sagesse se révèle être empiriquement crédible, elle pourrait « sauver » le modèle globaliste du caractère des critiques que nous avons analysées au fil du présent travail.

Différentes formes de sagesse

La sagesse est une forme de jugement supérieur capable de transformer le caractère et d'élever l'être humain à un niveau supérieur d'existence. C'est le noyau sémantique minimal de la notion. Y a-t-il plusieurs formes de sagesse ? Il n'y a pas de consensus sur cette question, qui, au fil de l'histoire, a été tranchée sur des bases plus « locales ». Nous verrons que la sagesse a une certaine tendance intrinsèque à être une notion unitaire : cela ne veut pas dire que les conceptions qui différencient plusieurs formes de sagesse sont fausses (en tout cas, nous n'avons pas l'espace pour évaluer une telle question), mais que les distinctions opérées sont généralement motivées par des raisons externes et indépendantes du concept de sagesse. Pour ainsi dire, la sagesse n'a pas de tendance « spontanée » à se diviser en plusieurs formes.

Les deux variables principales pour différencier plusieurs formes de sagesse sont l'origine de la sagesse (en tant que corpus de connaissances qui rendent possible le jugement sage) et le domaine d'application. En ce qui concerne l'origine, dans un contexte religieux il a été fréquent d'opérer une

distinction entre sagesse humaine et sagesse divine, où la deuxième serait supérieure à la première. D'après ces théories, les deux manifestent les propriétés distinctives de la sagesse, mais la sagesse divine issue d'une révélation est plus « performante » que la sagesse humaine, qui serait une forme amoindrie de sagesse. Nous n'allons pas nous intéresser à cette distinction, qui présuppose des thèses théistes. Mais il est clair qu'elle n'est pas motivée par des tensions internes à la notion de sagesse⁵⁴⁴.

Par contre, une distinction célèbre que nous ne pouvons pas ignorer est celle entre la sagesse appliquée au domaine pratique et la sagesse appliquée au domaine intellectuel et à l'enquête rationnelle. L'existence de deux domaines d'application différents n'est pas une raison suffisante pour conclure qu'on a affaire à deux formes de sagesse. Bien évidemment, l'enquête rationnelle et la vie pratique relèvent de deux domaines différents, mais c'est possible que la même forme de jugement sage soit mobilisée dans les deux cas.

Par définition, la sagesse est un savoir qui opère des arbitrages entre domaines : par exemple, une personne peut décider sagement de changer de travail, pour trouver un équilibre meilleur entre sa vie publique et sa vie privée, en vue d'une contribution plus efficace dans le long terme au bien être de sa famille, de sa communauté et de soi-même. Ce n'est pas évident qu'il faille postuler une sagesse spécifique à la gestion de la vie de famille et une forme de sagesse spécifique à la vie professionnelle : cette fragmentation serait une critique directe à la notion traditionnelle de sagesse. En ce qui concerne la sagesse épistémique et la sagesse pratique, la seule façon de défendre la distinction c'est de défendre une « autonomie modérée » du domaine de l'excellence intellectuelle (des vertus intellectuelles) du domaine de l'excellence pratique (les vertus morales). Si l'autonomie des vertus intellectuelles (c'est-à-dire l'ensemble de capacités et de facultés dont l'exercice explique l'excellence dans la vie intellectuelle) est totale, il est difficile de parler de « sagesse intellectuelle ». Comme nous l'avons vu, la notion de sagesse a une dimension holiste, il s'agit d'un savoir adapté à l'homme, globalement, aussi à la dimension pratique de sa vie. En ce sens, pour qu'il soit question de « sagesse intellectuelle » il faut forcément avoir une conception pragmatique de la connaissance : si nous faisons face à des choix théoriques, la sagesse intellectuelle nous pousse à réinterpréter la réalité d'une façon qui va se révéler globalement plus fructueuse à long terme, pour des êtres qui ont les limitations et le potentiel de l'homme. Par exemple, une préférence pour la simplicité des

⁵⁴⁴ D'un point de vue historique, les systèmes idéologiques et religieux semblent poser des contraintes externes et quelque peu « artificielles » à la notion de sagesse. Si la sagesse est une notion facilement reconnaissable dans un grand nombre de cultures, les conceptions qui soulignent son origine divine ou surnaturelle introduisent une dimension spécifique à la culture donnée. Voir Baltes, "Wisdom as Orchestration of Mind and Virtue", 73.

théories scientifiques peut être interprétée comme une valeur pragmatique favorisée par la sagesse intellectuelle : les réinterprétations les plus simples de la réalité seraient plus adaptées à l'homme. Dans la mesure où certains domaines de la connaissance n'ont pas de lien avec la dimension humaine⁵⁴⁵, la notion de sagesse intellectuelle n'est plus applicable. Par exemple, Aristote postule une vertu intellectuelle spécifique qui nous permet de saisir des réalités métaphysiques, éternelles et immuables (et complètement indépendantes de la dimension humaine) : c'est la « *sophia* ». Ce terme est souvent traduit par sagesse, mais en réalité il s'agit tout simplement d'une faculté cognitive fiable dans la saisie des réalités métaphysiques : les vérités métaphysiques nous sont dévoilées exactement telles qu'elles sont, mais dans ce processus il n'y a plus de trace des arbitrages propres à la sagesse intellectuelle.

« Pour Aristote, l'être et donc le domaine du connaissable est partagé en deux grands domaines : le nécessaire et le contingent. Comme il y a bijection entre disposition à connaître et objet connu, la division entre nécessaire et contingent entraîne une division dans les dispositions cognitives de l'homme : la capacité à connaître scientifiquement porte sur le nécessaire et la capacité à calculer ou à avoir des opinions vraies, sur le contingent, et plus précisément sur le bien relatif aux hommes. »

*« La sagesse (*sophia*) comprend deux dispositions qui épuisent le champ du savoir scientifique (...) le premier élément de la sagesse est la science (*epistémé*), c'est-à-dire la capacité à faire des raisonnements déductifs (...) Une autre disposition est nécessaire, qui permette de connaître de façon non démonstrative les principes : l'intellect (*nous*)⁵⁴⁶. »*

Si nous gardons la définition du noyau sémantique minimal de la notion de sagesse que nous avons donné, il apparaît que chez Aristote il y a une théorie de la sagesse pratique (prudence), et non pas de la sagesse intellectuelle (la *sophia* n'est pas une forme de sagesse), car Aristote a une conception totalement non pragmatiste de la connaissance intellectuelle des réalités ultimes et immuables.

D'autres théoriciens de la vertu intellectuelle effacent la notion de sagesse intellectuelle parce qu'ils sont « trop » pragmatistes : ils envisagent un lien trop fort entre vertus intellectuelles et vertus morales. C'est le cas de Zagzebski, qui défend l'unité entre psychologie épistémique et psychologie

⁵⁴⁵ Il faut entendre cela au sens de vérités éternelles et immuables qui ne concernent pas l'homme : certaines vérités éternelles peuvent concerner l'homme, par exemple, dans un cadre religieux, des connaissances théologiques sur la nature de Dieu.

⁵⁴⁶ Cyrille Bégorre-Bret, "Aristote : les vertus intellectuelles entre moralité et vérité," in *Le retour des vertus intellectuelles*, édition Thomas Bénatouïl et Michel Le Du (Les cahiers philosophiques de Strasbourg, 2006), 57-58.

morale⁵⁴⁷. D'après Zagzebski, le fonctionnement des vertus intellectuelles est structurellement homologue à celui des vertus morales. Il est clair qu'il faut concéder l'existence d'une différence minimale inévitable : trivialement, les vertus morales produisent des actions alors que les vertus épistémiques produisent des croyances. Mais, si on analyse une vertu intellectuelle de près, sur un plan plus substantiel et en gardant à l'esprit les caractéristiques essentielles d'une vertu morale, la ressemblance structurelle devient patente. Il n'y a aucune raison pour ne pas considérer une vertu intellectuelle comme une espèce particulière de vertu morale. Les vertus éthiques sont d'ailleurs très différentes entre elles : le courageux doit *toujours* maîtriser de très fortes émotions (la peur) alors que, par exemple, la justice n'est pas caractérisée par cette forme de « spécialisation » dans la gestion d'émotions très intenses. Si l'on met dans un même ensemble les vertus morales et les vertus intellectuelles, il n'y a aucune façon intéressante de faire le partage. *Les vertus intellectuelles prises globalement ne sont pas plus différenciées des vertus éthiques que les vertus éthiques ne le sont entre elles :*

- a. les vertus morales sont traditionnellement liées à un contrôle des passions (souvent intenses), mais il y a aussi des vertus intellectuelles qui doivent gérer des passions, par exemple la « passion pour la vérité ».
- b. Traditionnellement on insiste sur la possibilité de conflit entre vertus morales et vertus intellectuelles, mais deux vertus morales ne sont pas moins souvent en conflit entre elles qu'une vertu éthique et une vertu intellectuelle. Il se peut que, dans la même circonstance, on ne puisse pas être en même temps courageux et humble.
- c. Dans l'*Ethique à Nicomaque* (EN II.1.1103a14-20), Aristote affirme que les vertus intellectuelles sont des qualités qui peuvent être enseignées, alors que les vertus morales sont des habitudes qui sont acquises grâce à la pratique et l'entraînement. Cela n'est pas vrai. L'acquisition de bonnes habitudes intellectuelles requiert le même processus d'apprentissage progressif et en grande partie pratique que les vertus morales.
- d. Toujours d'après Aristote, « la vertu du caractère a affaire aux choses agréables et désagréables » (EE II.2.1220a36). Mais il y a des plaisirs typiquement associés à la recherche intellectuelle : la joie de faire une découverte, le plaisir de la confirmation d'une hypothèse, la satisfaction que donne une compréhension profonde de la réalité, etc.

Pour toute caractéristique prétendument distinctive, les vertus intellectuelles ne la possèdent moins que les vertus morales. Bref, d'après Zagzebski, la « psychologie épistémique » est substantiellement identique à la « psychologie morale » : la structure des deux genres de facultés est

⁵⁴⁷ Zagzebski, *Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge*.

identique, il s'agit donc en réalité d'un seul genre. Dans ce contexte, une seule forme de sagesse peut exercer son rôle de médiation et d'arbitrage sur toutes les vertus.

Ici nous n'avons pas vocation à explorer davantage ce débat, mais il est clair que la notion de sagesse intellectuelle est compatible uniquement avec une forme d'autonomie modérée des vertus intellectuelles par rapport aux vertus morales. Dans ce cas de figure, les vertus intellectuelles auraient besoin d'une coordination spécifique (la sagesse intellectuelle), tout en restant en contact avec les vertus morales (coordonnées par la sagesse pratique), de la même manière que deux groupes qui opèrent en coordination ont deux chefs. Si les vertus intellectuelles sont complètement isolées des vertus morales et de la dimension pratique de l'existence, elles deviennent de simples capacités cognitives fiables qui nous donnent accès à la vérité, sans que les arbitrages propres à la sagesse intellectuelle soient nécessaires. Si les vertus intellectuelles et les vertus morales sont complètement unifiées, il n'y a pas de raison de postuler deux formes de sagesse.

Encore une fois, d'un point de vue purement stratégique, nous pouvons remarquer que la sagesse a tendance à être une notion unitaire. La distinction entre plusieurs types de sagesse n'est cohérente que dans des contextes théoriques déterminés, et si on accepte d'autres thèses indépendantes (p.ex. l'existence d'une révélation divine, une certaine relation entre vertus intellectuelle et vertus morales, ainsi qu'une certaine conception pragmatique modérée du savoir).

Études empiriques du concept de sagesse du sens commun

Pour terminer notre brève introduction au concept de sagesse, il faut montrer qu'il existe bien un concept de ce type dans le sens commun. Comme nous l'avons déjà expliqué, les théories de la vertu s'inscrivent dans la continuité de concepts qui occupent une place centrale dans la vie de tous les jours. Bien sûr, les philosophes ou les doctrines des grandes traditions religieuses proposent des théories du courage, de la justice ou de la sagesse différentes et plus élaborées que celles du sens commun. Mais il faut au moins pouvoir justifier qu'il s'agit de théories portant sur les *mêmes* phénomènes. Sans quoi, la théorie serait complètement déconnectée des pratiques quotidiennes dans lesquelles on reconnaît des individus sages, on recherche leur avis, on se demande si notre décision a été suffisamment sage, etc.

Suite à notre analyse des concepts de traits chez le sens commun, nous savons bien que la reconstruction des idées du sens commun que les théories philosophiques prennent comme point de départ n'est pas forcément fiable. Par exemple, dans les explications des gens, la notion de trait occupe une place beaucoup moins importante que deux millénaires de théories philosophiques (et

un siècle de théories psychologiques) des traits ne le suggèrent. Ainsi, il est indispensable d'étudier empiriquement le concept ordinaire de sagesse: des surprises sont tout à fait possibles.

Heureusement, ces études ont été faites et le résultat est satisfaisant. Cette fois, l'image que la tradition s'est faite du sens commun correspond à la réalité. Avant la moitié des années 70, il n'y avait pas eu d'études empiriques sur la sagesse: l'étude du concept ordinaire de sagesse a inauguré le domaine. La méthodologie de ces études est très classique, dans le cadre de la psychologie de la personnalité. Il faut montrer deux choses: (1) un concept associé à la notion de sagesse existe, c'est-à-dire que le terme « sage » n'est pas utilisé comme un simple synonyme d'autres notions, ou alors de façon totalement ambiguë et incohérente; (2) ce concept se rapproche de la notion de sagesse que les psychologues retiennent comme hypothèse de travail, autrement dit, il est « reconnaissable » en tant que concept de sagesse. Il s'agit d'un processus typique de validation d'une notion: la première phase correspond à la fiabilité, la deuxième à la validité. Le résultat final est la preuve de l'existence d'un concept distinct des autres concepts ordinaires (notamment, dans l'usage, la sagesse ne doit pas être confondue avec l'intelligence ou la créativité) et qui se comporte de la manière souhaitée.

En ce qui concerne la fiabilité, des études ont montré que le terme « sagesse » n'est pas utilisé de la même façon que les termes « intelligence » ou « créativité ». Les gens l'emploient facilement: ils sont en mesure de trouver un grand nombre d'exemples de jugements sages ou de personnes sages⁵⁴⁸. Ce n'est pas comme si on les forçait à utiliser un concept qui n'appartient pas à leur expérience quotidienne. En plus, les caractéristiques qui lui sont associées sont les mêmes dans les tâches de production spontanée (on demande aux participants de décrire librement l'attitude d'une personne sage) et dans les tâches de choix forcés (on travaille avec une liste prédéterminée d'attributs).

Cohérence dans l'usage

Clayton⁵⁴⁹ a fait évaluer la proximité sémantique entre le terme « sage » et d'autres termes tels que « intelligent », « introspectif », « emphatique », « intuitif », etc. Les participants devaient quantifier le niveau de proximité entre un ensemble de ces termes et « sage », évaluant tous les couples possibles (par exemple: « sage-intelligent », « sage-intuitif », « intelligent-intuitif », etc.). De cette façon, on assistait à l'émergence de trois groupes dans chacun duquel les termes étaient en même

⁵⁴⁸ D. Sowarka, "Weisheit und weise Personen: Common-Sense-Konzepte alterer Menschen [Wisdom and wise persons: Common sense concepts in older people]," *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 21 (1989).

⁵⁴⁹ V. Clayton, "Erikson's theory of human development as it applies to the aged: wisdom as a contradictory cognition," *Human Development* 18 (1975).

temps très proches entre eux et avec le terme « sage ». Or, ces groupes étaient à peu près les mêmes indépendamment de l'âge et du genre des participants. Cela signifie que le concept de sagesse est utilisé de façon cohérente par plusieurs sous-groupes de la population. Dans une autre expérience, Sowarka⁵⁵⁰ a analysé les descriptions d'individus sages que chaque participant avait eu l'occasion de rencontrer personnellement dans sa vie. Dans le texte de ces descriptions libres d'individus que les participants avaient eux-mêmes choisis, Sowarka a mis en évidence certains attributs qui avaient tendance à être corrélés. On retrouve encore une fois l'émergence de trois ou quatre groupes d'attributs, à peu près les mêmes que dans l'expérience de Clayton.

La cohérence de la structure sémantique du concept de sagesse a été testée de façon encore plus systématique par Sternberg⁵⁵¹. Il a demandé à des participants d'écrire trois listes d'attributs, une qui caractérise le prototype de l'individu sage, une qui caractérise le prototype de l'individu intelligent et la dernière qui caractérise le prototype de l'individu créatif. Ensuite, il a demandé à d'autres participants de regrouper en autant d'ensembles qu'ils le souhaitent tous les attributs de la première expérience confondus. Les corrélations entre ces attributs dans le cadre d'une autoévaluation ont été également étudiées, ainsi que l'évaluation de lettres de recommandation spécialement conçues pour exprimer un certain nombre d'attributs puisés dans la liste de départ. Dans tous ces cas, la structure sémantique du concept de sagesse est identique et différente des structures sémantiques des concepts d'intelligence et de créativité. Qu'on s'exprime spontanément ou qu'on décide comment traiter une liste d'attributs déjà déterminés, on mobilise les mêmes dimensions. Le concept de sagesse est utilisé de façon remarquablement cohérente et homogène par à peu près tout le monde.

Du point de vue de la psychologie de la personnalité, cette découverte n'est pas surprenante. On pourrait remplacer « sage » par (presque) n'importe quel concept de trait, et les résultats seraient comparables. Comme nous l'avons montré dans le chapitre sur la théorie des traits du sens commun, dans la population adulte (la maîtrise est très tardive) les concepts de traits sont utilisés de façon uniforme et cohérente. Des personnes de niveau socio-économique différent et qui ne se connaissent pas évaluent de façon identique une connaissance commune, et la structure sémantique mise en évidence dans des tâches libres est toujours très proche de celle qui émerge dans les tâches à choix forcé. Cela va à l'encontre d'une conception répandue du sens commun selon laquelle les concepts

⁵⁵⁰ Sowarka, "Weisheit und weise Personen: Common-Sense-Konzepte alterer Menschen [Wisdom and wise persons: Common sense concepts in older people].", D. Sowarka et al., "The prototypical meaning of wisdom and wise persons," non publié (1990).

⁵⁵¹ R. J. Sternberg, "Implicit theories of intelligence, creativity, and wisdom," *Journal of personality and social psychology* 49, no. 3 (1985).

ordinaires seraient toujours vagues et ambigus. Dans la cognition ordinaire, l'association entre certains mots et certains concepts est sans doute souple et contextuelle: si on décrit le choix d'un étudiant âgé de dix ans comme étant sage, il s'agit bien évidemment d'un sens relâché de « sagesse », on ne s'engage pas véritablement à soutenir que l'enfant est sage. La possibilité de mobiliser les concepts de façon relâchée est une marque distinctive du sens commun. Dans certains contextes, un sens relâché de « sagesse » peut se rapprocher considérablement de la signification standard d'intelligence ou d'autres notions apparentées dans le schéma conceptuel du sens commun. Si on se réfère à ce type de phénomène, on peut bien sûr affirmer qu'il y a plusieurs sens de sagesse et que le sens commun n'établit que des distinctions floues et instables entre ses notions cruciales. Mais lorsque le terme « sage » est employé de façon « sérieuse » (dans un jargon technique, on dirait « ontologiquement engagé », voir le chapitre sur la théorie des traits du sens commun, section « anti-réalisme global »), on assiste à l'émergence d'une homogénéité remarquable.

Validité

Mais quel est le contenu de ce concept de sagesse, dont la structure sémantique de base semble être universelle? La cohérence dans l'usage d'un terme prouve la fiabilité, encore faut-il démontrer la validité, c'est-à-dire le fait que le concept ordinaire associé au terme « sagesse » est suffisamment semblable à la notion de sagesse qui a occupé l'attention des philosophes et des psychologues.

La validité n'est pas non plus très difficile à prouver. Dans les groupes d'attributs ou dimensions sémantiques mises en évidence par les expériences, on retrouve facilement des thèmes chers à la tradition. Par exemple, dans l'expérience de Clayton les dimensions sont: (1) un composant réflexif, formé par des qualités d'introspection; (2) un composant affectif, formé par des qualités de compréhension, d'empathie, de gentillesse et par une attitude pacifique; (3) un composant cognitif qui couvre des connaissances pratiques ainsi que l'expérience de vie. On retrouve très clairement les thèmes de l'attitude bienveillante du sage (trait de personnalité typique chez les individus sages, dans toute tradition), de sa compréhension profonde de soi-même et de sa compréhension des relations interpersonnelles. Ce dernier facteur est ultérieurement souligné par les exemples de sagesse donnés plus fréquemment par les gens. Comme l'analyse des interviews proposée par Sawarka le montre, les « actions sages » sont surtout des arbitrages dans des choix et des situations interpersonnelles particulièrement difficiles. Pour le sens commun, le sage est un très bon conseiller dans le domaine des relations humaines. Les autres études que nous avons citées dressent le même

tableau: une intelligence sociale très poussée ainsi que d'autres capacités nécessaires pour un bon jugement⁵⁵², des capacités d'analyse de soi-même et une aptitude à la bienveillance.

Variation dans l'usage

Quelles sont les sources de variation dans l'usage commun du concept de sagesse? Après avoir éliminé la variation qui relève d'un usage relâché du concept, dans le cadre d'un usage « sérieux » on remarque deux tendances. Les personnes âgées ont un concept de sagesse plus riche et articulé, les dimensions affective, cognitive et réflexive émergent plus clairement (les corrélations entre les composants de chaque dimension sont plus élevées)⁵⁵³. L'explication probable est qu'une tendance à être plus sage (associée statistiquement à l'âge) stimule la réflexion sur le concept de sagesse.

En ce qui concerne la deuxième tendance, des catégories professionnelles différentes soulignent des dimensions non équivalentes de la notion de sagesse. Notamment, les philosophes professionnels américains semblent avoir une notion de sagesse beaucoup plus proche de la notion d'intelligence que le reste de la population⁵⁵⁴. Ce type de variation est très représentatif du fonctionnement de la cognition ordinaire et de l'interaction entre les théories et le sens commun. Les concepts ordinaires ont souvent plusieurs dimensions, et les théories philosophiques ou psychologiques qui s'en inspirent finissent inévitablement par en souligner certains aux dépens des autres. Toutefois, dans le sens commun cet effet est limité: les dimensions qui sont mises en avant n'effacent pas complètement les autres et la structure sémantique de base reste toujours claire et reconnaissable. Même si on souligne le rôle du composant cognitif (l'intelligence), les composants réflexif (capacité d'auto-analyse) et affectif (bienveillance) ne sont jamais oubliés et émergent clairement dans l'analyse de la structure sémantique.

Section II évaluation du bien fondé empirique du rôle unificateur de la sagesse

Approches empiriques à la sagesse

On pourrait s'attendre au pire de la rencontre entre la réflexion sur la sagesse – dans le sens commun et les grandes traditions philosophiques et littéraires du monde – et la psychologie

⁵⁵² Sowarka, "Weisheit und weise Personen: Common-Sense-Konzepte alterer Menschen [Wisdom and wise persons: Common sense concepts in older people]."

⁵⁵³ Clayton, "Erikson's theory of human development as it applies to the aged: wisdom as a contradictory cognition."

⁵⁵⁴ Sternberg, "Implicit theories of intelligence, creativity, and wisdom."

empirique contemporaine. Vraisemblablement – pourrait-on penser – le niveau de nos connaissances scientifiques n'est pas suffisant pour juger directement de la possibilité d'un tel savoir, et les modèles scientifiques du fonctionnement de l'esprit risquent d'être étrangers au vocabulaire psychologique de la sagesse : un terrain de confrontation pourrait être difficile à trouver. Heureusement, la situation est meilleure. Si l'étude empirique de la vertu et du fonctionnement excellent de la personnalité n'en est effectivement qu'à ses balbutiements, le dialogue avec la psychologie empirique a déjà commencé. La psychologie empirique s'est efforcée d'étudier des notions proches de celle de sagesse depuis longtemps et, au moins depuis deux ou trois décennies, elle a essayé d'étudier directement la sagesse, aussi bien en tant que jugement sage qu'en tant que personnalité du sage.

Dans la suite de ce chapitre, nous allons d'un côté simplement exposer une série de travaux empiriques sur la sagesse : il s'agit d'efforts qui méritent d'être portés à l'attention des philosophes. Nous n'allons pas prendre position en faveur d'une approche plutôt que d'une autre, il s'agira seulement d'utiliser globalement ces travaux pour évaluer le bien-fondé empirique du rôle de la sagesse en tant que facteur d'unification des vertus et d'une personnalité harmonieuse et cohérente. Il va sans dire que le niveau de certitude des conclusions auxquelles nous souscrirons est largement *en deçà* de tout doute raisonnable. Il ne s'agit que d'opérationnalisations de la notion de sagesse : c'est n'est que le début d'un dialogue entre psychologie et philosophie. En même temps, on ne saurait pas ignorer ces conclusions uniquement parce qu'elles sont provisoires : il s'agit des seules traditions de psychologie empirique qui soient jamais intéressées à la notion de sagesse.

Baltes

Le programme de recherche le plus avancé sur la sagesse en tant que jugement est celui qui a été dirigé pendant les deux dernières décennies par feu Baltes (1939-2006) à l'Institut Max Plank pour le développement humain de Berlin⁵⁵⁵. Baltes était un spécialiste du vieillissement cognitif et s'intéressait au rôle de la sagesse dans un vieillissement heureux. Son intuition fondamentale a été de mobiliser la littérature sur la différence entre l'intelligence « fluide » et « cristallisée » pour comprendre la nature bipolaire de la sagesse. L'intelligence fluide est une capacité à résoudre des problèmes qui ne se base pas sur des connaissances déjà possédées : un enfant qui, ne connaissant rien à la géométrie, arrive à résoudre facilement les premiers exercices qu'il rencontre, fait preuve

⁵⁵⁵ P. B. Baltes et U. M. Staudinger, "The Search for a Psychology of Wisdom," *Current Directions in Psychological Science* 2, no. 3 (1993).

d'intelligence fluide. Cette notion correspond à l'usage du sens commun de la notion de l'intelligence : intuitivement, on ne dirait pas que quelqu'un est très intelligent s'il va résoudre un problème grâce à une grande quantité de connaissance. Il y a cependant un piège dans le choix du label « fluide » : l'intelligence est fluide par rapport à l'intelligence cristallisée parce qu'elle n'est pas liée à un corpus d'expériences et de connaissances stockées...mais, paradoxalement, cela signifie qu'elle dépend des qualités cognitives de base du système (dans une métaphore informatique, elle dépend du hardware), elle est donc plus « fixe » que l'intelligence cristallisée !

L'intelligence cristallisée dépend de l'expérience, des habitudes et des styles cognitifs, ainsi que des connaissances accumulées. Avoir résolu des problèmes semblables par le passé, avoir développé des stratégies cognitives efficaces ou tout simplement avoir beaucoup de connaissances sont autant de conditions qui font augmenter considérablement les performances intellectuelles, par exemple en tant que mesurées dans un test de QI. Pour cette raison, le QI monte régulièrement avec l'âge et l'expérience. La littérature contemporaine sur l'expertise et l'excellence ne cesse de revoir à la hausse le rôle de l'intelligence « cristallisée » et accumulée avec l'expérience, par rapport à la performance « de base » du système⁵⁵⁶. Encore une fois, il y a un piège linguistique : l'intelligence est cristallisée parce qu'elle dépend (à un moment donné) d'un stock de connaissances accumulées, mais en fait l'intelligence cristallisée n'a rien de fixe, car elle augmente rapidement avec l'expérience !

Quelle est la relation entre le phénomène de la plasticité cérébrale et l'intelligence fluide ? Le cerveau répond à l'entraînement massif par des modifications de sa structure neurobiologique qui peuvent être profondes. L'entraînement intensif pendant plusieurs années génère, comme nous l'avons vu dans le chapitre précédent, un véritable module cognitif « acquis », dans le domaine en question.

Le maître aux échecs a une mémoire prodigieuse des positions de jeu, mais cette prestation ne s'étend pas au-delà du domaine : l'extension de la mémoire de travail et de long terme est spécifique aux échecs, si bien que le joueur garde par ailleurs ses propres limitations de base ou « hardware » (qui sont à la base de l'intelligence fluide). Dans les termes employés par Baltes, il y a une différence entre la mécanique (dépend du hardware : intelligence fluide) et la pragmatique (intelligence cristallisée issue de l'expérience) cognitives. Alors que la mécanique cognitive décline progressivement avec l'âge, les performances augmentent grâce à l'accumulation d'intelligence

⁵⁵⁶ M. E. Beier et P. L. Ackerman, "Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment," *Psychology and aging* 20, no. 2 (2005), Ericsson, "The influence of experience and deliberate practice on the development of superior expert performance."

cristallisée, jusqu'au point où la perte en fonctionnalités basiques devient trop importante pour que tout le système n'en soit pas affecté.

Baltes pensait que la notion d'intelligence cristallisée nous donne un excellent point de départ pour comprendre la sagesse en tant que savoir. Dans cette littérature psychologique, on retrouve la polarité entre d'un côté le savoir affirmatif et positif et de l'autre le déploiement d'heuristiques. L'intelligence cristallisée est rendue possible par un processus d'accumulation de connaissances et d'expériences bien déterminées : ce n'est pas une notion qui relève entièrement de la mise en perspective. Si on prend l'exemple des mathématiques, au départ il y a bien la connaissance d'un grand nombre de résultats et théorèmes, ainsi que l'expérience dans la résolution d'exercices. Ce savoir substantiel alimente des compétences heuristiques et la possibilité de s'orienter intelligemment lorsqu'on est confronté à des problèmes nouveaux. Comme dans le cas de la sagesse, la dimension heuristique de l'intelligence cristallisée ne relève pas simplement d'une mise en perspective dialectique (voire sceptique), déconnectée de toute contrainte de succès. Au contraire, par définition l'intelligence cristallisée rend extrêmement efficace dans des tâches cognitives exigeantes.

D'après Baltes, la sagesse est une forme spécifique d'intelligence cristallisée (ou pragmatique cognitive) portant sur les questions existentielles les plus profondes. Son programme d'étude rigoureux représente l'un des meilleurs exemples d'interdisciplinarité entre philosophie et psychologie sur le thème de la vertu. Ses expériences testent la capacité de jugement des sujets par rapport à des situations et des choix difficiles. Un exemple de cas hypothétique présenté est le suivant :

« Joyce, une veuve sur la soixantaine, a obtenu récemment un diplôme en management et a créé sa propre entreprise. Elle attendait avec enthousiasme ce nouveau défi. Néanmoins, elle vient de savoir que son fils a été laissé seul avec deux enfants en charge. Joyce est en train de prendre en considération les options suivantes : elle pourrait abandonner son activité et vivre avec son fils ou continuer mais prendre en charge financièrement le suivi des deux enfants. Formule un plan qui détaille ce que Joyce pourrait faire dans une période de 3 à 5 ans. De quel genre d'informations supplémentaire y aurait-il besoin ⁵⁵⁷? »

Les réponses sont enregistrées après que les participants ont été entraînés à penser « à haute voix », une technique typique de la psychologie de l'expertise, censée révéler le plus possible sur les styles

⁵⁵⁷ P. B. Baltes et al., "People Nominated as Wise: A Comparative Study of Wisdom-Related Knowledge," *Psychology and Aging* 10, no. 2 (1995): 159.

cognitifs et les stratégies de raisonnement des experts. Toute réponse est codée sur des échelles de 1 à 7, par rapport à cinq critères pertinents pour la sagesse :

(1) connaissance factuelle riche sur des sujets tels que les différentes phases de la vie humaine, les trajectoires individuelles possibles pour traverser ces phases, les relations interpersonnelles, les normes sociales, les événements cruciaux dans la vie, les compromis possibles entre le bien-être individuel et collectif, etc. Ce critère représente la sagesse en tant que savoir substantiel sur les aspects le plus importants de la vie humaine : la sagesse n'est pas un catéchisme moral pédant et détaillé, toutefois, un noyau de savoir substantiel est un composant indispensable de sa structure bipolaire.

(2) connaissance procédurale sur des stratégies de gestion d'événements décisifs dans la vie

(3) contextualisme relatif aux différentes phases de la vie : la capacité à mettre en perspective les exigences et les désirs d'une personne par rapport à son âge, contexte culturel et social

(4) sensibilité au relativisme des valeurs de différentes cultures, compatible avec une attention à un noyau de valeurs universelles

(5) conscience et capacité de gestion du niveau d'incertitude inhérent à la vie humaine.

Une personne est considérée comme sage (au sens du jugement) si elle obtient une note de cinq ou plus dans la plupart de ces critères.

Baltes a testé des gens de tout âge (de l'adolescence à 89 ans) et milieu social dans ce paradigme expérimental. Contrairement aux attentes, l'âge n'est pas corrélé étroitement avec la sagesse de jugement. La seule corrélation intéressante se retrouve parmi les adolescents et les jeune adultes : comme on pourrait facilement l'imaginer, de 14 à 25 ans la sagesse de jugement augmente de façon considérable⁵⁵⁸. Mais après, les gains apportés par l'âge sont modestes. Cela va dans le sens de la psychologie de l'expertise contemporaine : qui vit longtemps a davantage d'expérience de vie, mais, passé un certain niveau, la simple « pratique » n'est pas suffisante pour une augmentation des performances, en absence d'un « entraînement délibéré » et systématique. L'âge est une condition nécessaire mais pas suffisante pour dépasser les niveaux de sagesse possibles à 25 ans.

En ce qui concerne le milieu professionnel, les métiers centrés sur le conseil et la compréhension de la situation personnelle d'autrui ont un avantage sur les autres. En particulier, les psychothérapeutes atteignent les scores les plus élevés parmi toutes les professions, leurs notes étant aussi bonnes que celle d'une série de personnalités publiques allemandes nominée en tant que « spécialement sages »

⁵⁵⁸ M. Pasupathi, U. M. Staudinger, and P. B. Baltes, "Seeds of Wisdom: Adolescents' Knowledge and Judgment About Difficult Life Problems," *Developmental Psychology* 37, no. 3 (2001).

par des journalistes⁵⁵⁹. Parmi les psychothérapeutes, on retrouve le phénomène d'absence d'amélioration liée à l'âge prédit par la psychologie de l'expertise (contre le sens commun) : tout comme les médecins plus expérimentés ne sont pas forcément meilleurs que les jeunes (sauf dans un champ de spécialisation très restreint) les psychothérapeutes plus âgés ne manifestent pas un meilleur jugement que leur collègues plus jeunes.

En ce qui concerne d'autres corrélats, la sagesse de jugement ne semble pas être associée de façon significative à des mesures d'intelligence. Par contre, certains styles cognitifs et la créativité sont les indices les plus importants, avec l'ouverture à l'expérience (dimension du *big five*) en deuxième place. La sagesse de jugement est aussi corrélée de façon positive à des stratégies coopératives de gestion de conflit et de façon négative à des stratégies basées sur la domination ou la soumission. Cependant, et contrairement à ce à quoi on pourrait s'attendre, elle n'est pas liée de façon cohérente avec le névrotisme et la complaisance du *big five*⁵⁶⁰.

Comment interpréter ces résultats ? La conclusion la plus importante est la suivante : les mesures élaborées par Baltes ont été validées de façon plutôt satisfaisante, si bien qu'on possède un indicateur fiable du genre de savoir auquel la tradition s'est toujours référée. En même temps, la sagesse ainsi mesurée ne semble pas être corrélée aux traits distinctifs de la sagesse en tant que personnalité vertueuse, d'où la conclusion décisive : être sage au sens de l'intériorisation d'un savoir portant sur l'art de vivre n'est pas suffisant pour assurer le développement harmonieux de la personnalité. Cela signifie que la psychologie contemporaine nous donne des raisons de douter du rôle assigné à la sagesse par la théorie de la vertu.

Revenons sur les prémisses qui nous conduisent à cette conclusion. D'abord, la validation des mesures de Baltes. En psychologie, la validation consiste à montrer que des tests mesurent bien la variable qu'ils visent, au lieu d'autre chose. Un test est fiable s'il mesure un variable réelle au lieu de produire des résultats totalement inconsistants : on juge la fiabilité par le fait que les mêmes personnes obtiennent des résultats proches lorsqu'ils repassent le test dans plusieurs occasions (fiabilité test-retest). Mais un test fiable peut mesurer autre chose par rapport à ce qui était visé. Dans la phase de validation, on cherche à établir si la variable mesurée se comporte comme on s'attendrait, dans l'hypothèse qu'elle est la variable visée.

Expliqué de cette façon, le processus de validation pourrait paraître absurde. Car il semble qu'un test est d'autant plus valide qu'il ne déçoit pas nos attentes. Cela serait en même temps très

⁵⁵⁹ Baltes et al., "People Nominated as Wise: A Comparative Study of Wisdom-Related Knowledge."

⁵⁶⁰ U. Kunzmann and P. B. Baltes, "Beyond the Traditional Scope of Intelligence: Wisdom in Action," *Models of intelligence: International perspectives* (2003).

conservateur et inutile : si tout ce que nous savons déjà est correct, il n'y a pas besoin d'étudier un phénomène. Mais cette critique est déplacée. D'un côté, aucune variable ne se comporte exactement comme on l'attendait : dans la mesure où elle le fait jusqu'à un niveau acceptable, cela nous permet de l'identifier. Dans ce cas de figure, toute surprise ultérieure correspond à une découverte. De l'autre côté, la cohérence des résultats n'est pas jugée uniquement par rapport au sens commun et à la notion naïve de départ. Si l'on cherche à valider un test censé mesurer le niveau de courage, il ne s'agit pas de rechercher un maximum de conformité avec la conception du courage du sens commun, mais plutôt une cohérence maximale avec toutes les données pertinentes que nous avons, tout ce que nous savons sur le sujet.

Ces remarques sont importantes dans le cas de la sagesse. Si on retrouve parfaitement la polarité entre connaissance factuelle-substantielle (relations interpersonnelles, normes sociales, événements cruciaux dans la vie etc.) et heuristique (sensibilité au relativisme des valeurs, capacité de gestion du niveau d'incertitude inhérent à la vie humaine, etc.) qui caractérise la notion traditionnelle de sagesse, on découvre que, contrairement à ce que le sens commun pense, la sagesse n'augmente pas forcément avec l'âge. Cela est cohérent avec ce que nous savons de la psychologie de l'excellence : l'âge et la pratique ne sont pas suffisants sans un entraînement délibéré et systématique. Le rôle de l'entraînement est souligné davantage par le fait que les psychothérapeutes (experts dans l'activité de comprendre et conseiller les gens) obtiennent des résultats meilleurs que n'importe quelle autre catégorie et ils se placent au même niveau que des personnalités publiques spécialement nommées pour leur sagesse. C'est un autre thème que l'on ne retrouve pas dans la conception du sens commun mais qui, loin d'invalider le test développé par Baltes, lui confère davantage de crédibilité, étant donné qu'il est cohérent avec notre vision globale de l'expertise⁵⁶¹.

⁵⁶¹ Si on cherche des aspects problématiques dans les études de Baltes, on peut souligner le fait que les études avec des personnes d'âges différents ne sont pas longitudinales, c'est-à-dire qu'elles ne mesurent pas le niveau de sagesse des mêmes personnes à distance de quelques décennies. Des différences inter-générationnelles de milieu culturel et de formation reçue peuvent affecter les résultats. Notamment, il est possible qu'il soit plus facile pour les jeunes d'acquérir les connaissances pertinentes pour la sagesse, à cause de la plus grande exposition au matériel culturel pertinent dans la société contemporaine. Il faudrait aussi trouver une façon de corriger l'avantage que chaque génération a dans le traitement des problèmes relatifs à son âge. Un autre problème est l'usage de la procédure de « nomination » pour désigner les sages. Dans la littérature sur l'expertise, le vrai niveau de compétence des « nommés » s'est montré souvent douteux. Heureusement, la validation des tests de Baltes ne dépend pas de manière cruciale des résultats des nommés, car la prestation moyenne dans de différentes professions et milieux sociaux respecte les prédictions (ex : les psychothérapeutes sont plus sages), indépendamment de la nomination. Pour un compte rendu critique de l'approche de Baltes, voir M. Ardelt, "Wisdom as Expert Knowledge System: A

Sternberg

Une autre approche psychologique remarquable à la sagesse est celle de Sternberg. Son domaine d'expertise n'est pas la gérontologie et l'étude du vieillissement sain et heureux, mais l'étude de l'intelligence. La mobilisation de la littérature sur l'intelligence est une constante dans le nouveau domaine de l'étude psychologique de la sagesse : les deux notions (intelligence et sagesse) sont relativement proches et l'intelligence a occupé l'attention des psychologues pendant tout le XX^e siècle. Balte s'est appuyé sur la distinction entre intelligence fluide et intelligence cristallisée, Sternberg est célèbre pour sa théorie tripartite de l'intelligence.

Les trois composants proposés sont les suivants : intelligence analytique, intelligence pratique et intelligence créative⁵⁶². L'intelligence analytique est la forme la plus étudiée : il s'agit de l'intelligence mesurée par le QI et associée au succès scolaire et académique. Il paraît que cette intelligence n'est pas strictement associée au succès non académique, notamment dans le monde du travail : les explications profondes de ce résultat sont nombreuses, mais on s'accorde à postuler l'existence d'une (ou plusieurs) forme(s) d'intelligence pratique, responsable des performances dans la vie réelle. Enfin, l'intelligence créative serait déployée pour affronter avec succès des situations nouvelles, auxquelles on n'est pas préparé.

Sternberg associe la sagesse à une forme d'intelligence pratique. Le choix est évident et s'inscrit clairement dans le modèle général de la vertu en tant qu'excellence pratique. « *L'intelligence pratique est la capacité à appliquer à l'expérience plusieurs ressources du traitement de l'information typique de l'intelligence, dans le but de s'adapter, de façonner ou de sélectionner les environnements* »⁵⁶³. La notion d'intelligence pratique telle qu'elle est conçue par Sternberg est également proche de l'intelligence cristallisée. Comme dans le cas de cette dernière, son fonctionnement dépend d'une base de connaissances, dont la mobilisation permet de gérer de situation suffisamment similaires ou au moins pertinentes, quoique nouvelles. Cette base est constituée par des connaissances tacites. La quantité de connaissances tacites pertinentes est le facteur qui détermine de la façon la plus significative les performances de l'intelligence pratique dans un domaine donné.

Critical Review of a Contemporary Operationalization of an Ancient Concept," *Human Development* 47, no. 5 (2004).

⁵⁶² Sternberg, "Implicit theories of intelligence, creativity, and wisdom."

⁵⁶³ R. J. Sternberg, "A Balance Theory of Wisdom," *Review Of General Psychology* 2 (1998): 352.

Il existe plusieurs méthodes pour mesurer le niveau de connaissance tacite d'experts ou de débutants. Il s'agit de variantes des méthodes classiques de la psychologie de l'expertise qu'on a déjà évoquées dans le chapitre précédent, notamment des interviews où l'expert est amené à « penser à haute voix » et justifier ses choix. Dans ce cadre expérimental, il est possible de pousser l'expert à faire une référence (au moins indirecte) aux structures cognitives qui sont censées encoder la connaissance tacite : notamment scripts, prototypes ou exemplaires de situations d'actions. À un niveau de généralité qui reste en deçà de tout désaccord dans la littérature contemporaine, on peut affirmer que la performance pratique est rendue possible par l'activation d'une base de données très vaste, constituée par des cas et des situations typiques pertinentes pour le contexte d'action. L'accès aux bonnes données est assuré par la perception (inconsciente) d'indices dans la situation présente. Ainsi, par analogie, la situation est immédiatement associée à un cas précis dans le répertoire. Parmi les psychologues, le désaccord réside dans la nature exacte des structures psychologiques qui composent ce vaste répertoire, résultat d'années d'expérience. En pensant à haute voix, l'expert peut révéler indirectement ses prototypes, scripts ou exemplaires de référence. La complexité et la variété de ces structures cognitives (autrement dit, les dimensions et la complexité de la base de données) peuvent être mesurées. Un maître de tennis peut faire référence à des centaines ou à des milliers de situations typiques et néanmoins distinctes qui peuvent se produire pendant un match de tennis. Un novice ne va évoquer qu'un nombre beaucoup plus limité de cas. Des mesures de ce type ont montré que, dans plusieurs professions (par exemple, le management ou la vente), la quantité de connaissance tacite est plus étroitement associée à la performance que l'intelligence analytique (QI)⁵⁶⁴.

D'après Sternberg, la sagesse est une forme spécifique d'intelligence pratique :

« La sagesse est définie comme l'application de la connaissance tacite, ayant en vue ses propres valeurs, pour atteindre le bien commun (a) par le biais d'un équilibre entre des multiples intérêts intra-personnels, interpersonnels et extra-personnels et (b) en trouvant un équilibre parmi les réponses possibles à des contextes environnementaux : adaptations à des environnements existants, façonnement des environnements existants et sélection de nouveaux environnements⁵⁶⁵. »

Plutôt que de type spécifique d'intelligence pratique, il s'agit d'un usage spécifique de l'intelligence pratique, car la sagesse n'est définie que par le but au service duquel on mobilise cette connaissance.

⁵⁶⁴ Ibid.

⁵⁶⁵ Ibid.: 353.

En ce sens, Sternberg fait de l'économie théorique : il n'est pas nécessaire de postuler une notion psychologique *sui generis*, que ce soit une forme de savoir ou une structure de personnalité. Il n'y a pas de nouvelles « espèces naturelles psychologiques » à introduire. Certes, chaque usage spécifique entraîne des adaptations : mobiliser un large ensemble de connaissances tacites pour faire des arbitrages complexes entre des intérêts individuels et collectifs est une compétence qui demande des années d'entraînement. Mais cela est vrai aussi d'un nombre indéfini d'autres usages possibles des mêmes connaissances : il est difficile aussi d'arbitrer pour maximiser son propre intérêt, ou celui de sa famille et de ses alliés. Un dictateur qui se maintient indéfiniment au pouvoir et un individu sage font preuve du même niveau de virtuosité pratique. D'un point de vue psychologique, la conception de la sagesse de Sternberg est « banalisante » : on ne fait plus face à une notion qui a une véritable spécificité psychologique. La sagesse relèverait d'une famille de notions qui désignent un usage excellent de l'intelligence pratique. Sternberg ne retient même pas l'idée que la sagesse s'exprime nécessairement sur un vaste ensemble de domaines :

« La théorie de l'équilibre suggère que la sagesse est au moins en partie spécifique au domaine, car la connaissance tacite est acquise à l'intérieur d'un contexte ou ensemble de contextes. Elle est typiquement acquise en sélectionnant de nouvelles informations qui sont pertinentes par rapport à ses propres motivations pour apprendre dans ce contexte, en comparant de façon sélective les nouvelles informations avec les vieilles données, pour voir comment les nouveautés s'insèrent dans le cadre des connaissances déjà acquises, et en combinant de façon sélective l'information pour chercher à créer un tout cohérent⁵⁶⁶. »

« Les individus qui acquièrent la sagesse dans un contexte peuvent bien être ceux qui seraient capables de la développer dans un autre contexte, mais la connaissance tacite dont on a besoin pour être sage dans des contextes différents peut varier⁵⁶⁷. »

On retrouve le thème de la spécialisation qu'on a déjà évoqué dans le chapitre sur l'expertise : un niveau d'excellence dans un domaine prédispose à (et peut-être entraîne) un certain niveau minimal dans des compétences proches. Mais en ce qui concerne l'excellence, on retrouve une spécialisation et une fragmentation totale : avant d'atteindre à nouveau un niveau d'excellence, un joueur d'échecs qui s'intéresse à d'autres jeux de table ou un musicien qui change d'instrument doivent passer par

⁵⁶⁶ Ibid.: 357.

⁵⁶⁷ Ibid.

un processus qui – s’il est effectivement moins long – reste comparable à celui d’un novice, en termes d’effort et de difficulté⁵⁶⁸.

Il est intéressant de souligner les différences entre l’approche de Sternberg et celui de Baltes : Sternberg ne conteste certainement pas que la sagesse soit caractérisée par les cinq critères opérationnels utilisés par Baltes dans ses expériences. En ce sens, pour opérer une médiation en vue du bien commun, à moyen et court terme, il est sans doute très utile de posséder : une connaissance factuelle riche sur des sujets tels que les différentes phases de la vie humaine; une connaissance procédurale sur des stratégies de gestion d’événements décisifs dans la vie; la capacité à mettre en perspective les exigences et les désirs d’une personne par rapport à son âge, contexte culturel et social; une sensibilité au relativisme des valeurs de différentes cultures, compatible avec une attention à un noyau de valeurs universelles. Néanmoins, Sternberg pense que cette liste est incomplète, car on ne peut pas énumérer de façon exhaustive les processus psychologiques et les compétences nécessaires pour atteindre le bien commun et opérer les arbitrages qui définissent la sagesse. On retrouve ici la « banalisation » psychologique de la sagesse qui est typique de son approche : la sagesse est définie par un but externe à l’individu, notamment la visée du bien commun, et elle mobilise des processus psychologiques qui ne sont pas spécifiques à la poursuite de ce but. La sagesse n’est pas une espèce naturelle psychologique. Selon le contexte, une grande variété de compétences et de processus psychologiques différents peut sous-tendre le fonctionnement de la sagesse : la liste ne saurait être exhaustive. En se référant aux théories de la sagesse en tant que personnalité, Sternberg écrit :

« La nature de l’équilibre dans la théorie proposée ici est, dans une certaine mesure, différente de la nature de l’équilibre dont il est question dans d’autres modèles, qui proposent un équilibre intra-personnel entre des systèmes cognitifs, conatifs et affectifs. Dans le modèle présenté, l’individu peut utiliser n’importe quel système interne pour formuler des jugements. (...) L’équilibre intra-personnel ne se réfère pas à un équilibre entre ces systèmes, mais entre les différents types d’intérêts qu’on possède⁵⁶⁹ »

Par rapport aux ambitions de la théorie de la vertu, aussi bien Baltes que Sternberg proposent une notion très modeste de sagesse en tant que jugement. Il est possible d’être plus sage dans un certain domaine que d’en d’autres (ce point est souligné davantage par Sternberg), et il est

⁵⁶⁸ K. A. Ericsson and P. Ward, “Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance,” *Current Directions in Psychological Science* 16, no. 6 (2007).

⁵⁶⁹ Sternberg, “A Balance Theory of Wisdom,” 359.

certainement possible d'avoir de conseils très sages à donner, alors qu'on ne les applique pas dans sa propre conduite. L'exemple le plus saillant est celui des thérapeutes, souvent capables d'aider les autres mais incapables de s'aider soi-même. L'idéal d'un savoir capable d'harmoniser totalement la personnalité humaine est très lointain.

Baltes et Sternberg concordent avec Aristote sur la nature métacognitive de la sagesse. En revanche, même si la sagesse est nécessaire pour résoudre des conflits entre des corpus de connaissances pratiques déterminées, on peut avoir cette capacité de résolution seulement dans certains domaines et pour certaines compétences.

L'argument d'Aristote selon lequel il ne peut y avoir de vertu complète et parfaite en absence d'une forme de sagesse globale et complète a deux points faibles. Suivant sa logique, il n'y aurait que deux possibilités: soit on possède un ensemble déconnecté de compétences pratiques indépendantes et autonomes, mais qui prennent en charge uniquement les problèmes pratiques « mono domaine » qui génèrent peu de conflits, soit on possède un ensemble parfaitement cohérent de compétences qui couvre de manière presque parfaite tous les domaines, grâce au rôle de médiation de la sagesse.

Première difficulté: la fonction métacognitive d'arbitrage de la sagesse ne doit pas forcément s'appliquer dans tous les domaines. Certes, une décision sage prend en considération toutes les valeurs: une décision issue d'un arbitrage entre courage et générosité ne peut être sage si, par ailleurs, elle est injuste. L'arbitrage a lieu parmi toutes les vertus pertinentes dans un contexte donné. Le but étant une action efficace et appropriée, une médiation partielle serait insuffisante et inutile. Or, le fait que la sagesse est par définition une capacité de médiation parmi *toutes* les vertus n'entraîne pas l'impossibilité de dissocier les domaines d'application. On peut être en mesure d'opérer une médiation parmi toutes les vertus dans la vie professionnelles, mais non dans la vie privée.

Deuxième difficulté: le lien indissociable entre vertu pleine et sagesse pourrait être valide uniquement à l'intérieur de chaque domaine. Autrement dit, soit on possède toutes les vertus et la sagesse dans la vie professionnelle (ou un autre domaine pertinent) soit on n'est ni vertueux (on ne possède aucune vertu) ni sage au travail. La sagesse et la vertu dans un domaine seraient compatibles avec le manque de sagesse et de vertu dans d'autres. Cette thèse a été défendue par Badhwar⁵⁷⁰.

La thèse de l'union indissociable « locale » entre sagesse et vertu est confrontée au problème suivant: l'intégration des vertus doit être totale seulement s'il s'agit de résoudre tous les arbitrages possibles. Mais, dans tout domaine pratique, il y a un certain nombre de problèmes qui sont

⁵⁷⁰ N. K. Badhwar, "The Limited Unity of Virtue," *Noûs* 30, no. 3 (1996).

beaucoup plus fréquents que d'autres: si on abandonne l'idéalisation d'une vertu impeccable, il est possible d'envisager un niveau de vertu et de sagesse importants mais compatibles avec certains points faibles et avec l'incapacité à gérer des problèmes rares. En ce sens, même à l'intérieur d'un domaine précis, la dichotomie aristotélicienne « absence de vertu/vertu complète » n'est pas valide. Elle le serait uniquement si on s'intéresse à une forme de vertu parfaite et sans faute, mais, d'un point de vue pratique, une vertu très développée, même si non parfaite, correspond à une vision réaliste de l'excellence.

Les approches empiriques à la notion de sagesse rejettent la thèse traditionnelle de l'unité des vertus assurée par le rôle de médiation de la sagesse. Pour compléter notre analyse, nous devons nous intéresser aux approches empiriques à la sagesse en tant que structure de personnalité.

Approches empirique à la sagesse en tant que personnalité

Depuis deux ou trois décennies, plusieurs programmes de recherche ont cherché à mesurer et étudier la sagesse en tant que personnalité vertueuse et harmonieuse. Il s'agit d'efforts moins innovants, qui ne font qu'appliquer l'approche des traits généraux à la sagesse, cherchant à développer et valider des questionnaires. Les résultats confirment le potentiel et les limites de la tradition des traits, que nous avons déjà analysée. En général, il est possible de valider avec succès des tests qui mesurent une tendance *générale* et *peu fiable* correspondant à la notion visée (voir chapitre sur la cohérence). La sagesse a été étudiée dans ce cadre.

L'auteur le plus représentatif de l'approche des traits à la sagesse est Ardel⁵⁷¹. Ardel adopte une méthodologie très pratique : élaborer un test qui mesure la sagesse à partir d'items et de subsections de tests qui existent déjà dans la littérature. En un siècle de travail, la psychologie de la personnalité a cherché à mesurer des centaines (voire des milliers) de traits différents : pour tout terme de trait présent dans le langage naturel, il existe des tests qui essaient de mesurer la variable psychologique correspondante (ou une variable qui s'en rapproche). Ainsi, si on postule l'existence de traits de haut niveau qui auraient des propriétés spécifiques (traits qui sont formés par la co-variation de traits de premier niveau, ou facettes), il est facile d'assembler un test, choisissant « à la carte » des facettes de tests déjà validés.

⁵⁷¹ Nous allons nous concentrer sur cette approche. Pour d'autres théories moins développées, voir W. A. Achenbaum et L. Orwoll, "Becoming wise: a psycho-gerontological interpretation of the Book of Job," *International Journal of Aging and Human Development* 32, no. 1 (1991).

Ardtel postule que la sagesse est un macro-trait qui aurait trois dimensions principales : un ensemble de compétences cognitives, un ensemble de compétences réflexives et un ensemble de compétences affectives. Cette schématisation reprend plusieurs des caractéristiques de la sagesse qu'on a déjà énumérées et qui sont présentes dans le sens commun ainsi que dans les traditions historiques.

Parmi les compétences cognitives, on retrouve : l'aptitude à comprendre une situation ou un phénomène en profondeur, la connaissance des aspects positifs et négatifs de la nature humaine, la reconnaissance de l'ambiguïté et de l'incertitude qui caractérisent la vie humaine, la capacité à prendre des décisions importantes et efficaces, malgré cette dimension d'imprévisibilité. Parmi les compétences réflexives, on peut énumérer : la capacité et la disposition à analyser les phénomènes et les événements depuis plusieurs points de vue, l'absence de projections injustifiées sur autrui à partir de sa propre situation subjective et émotionnelle. Les compétences affectives les plus importantes sont : la présence d'émotions et de comportements positifs envers autrui, l'absence d'émotions et de comportements indifférents ou négatifs envers autrui⁵⁷².

Le travail d'assemblage à partir des questionnaires déjà existants a permis le développement d'un questionnaire spécifique à la sagesse, le 3D-WS (trois dimensions de la sagesse, *wisdom*). Ce questionnaire est composé par 14 items pour la dimension cognitive⁵⁷³, 12 items pour la dimension réflexive⁵⁷⁴ et 13 items pour la dimension affective⁵⁷⁵. Ces items sont censés mesurer plusieurs facettes, dans chacune des trois dimensions : les facettes sont aux dimensions de haut niveau (cognition, réflexion, affectivité) ce que les facettes du *big five* sont à des super-traits tels que l'extraversion ou le névrotisme.

D'après Ardel, la sagesse réside dans l'équilibre entre les trois dimensions : il n'est pas possible d'acquérir une compréhension objective de la vie humaine, dans ses aspects positifs ou négatifs, si on n'a pas lutté avec succès contre les biais subjectifs qui nous poussent à interpréter la réalité à partir de nos convictions et de nos besoins. En ce sens, le travail réflexif sur soi-même est une étape

⁵⁷² Ardel, "Wisdom as Expert Knowledge System: A Critical Review of a Contemporary Operationalization of an Ancient Concept," 275.

⁵⁷³ Ex : « l'ignorance est une bénédiction » ; « les gens sont bons ou mauvais » ; « j'hésite à prendre des décisions importantes après y avoir réfléchi », etc.

⁵⁷⁴ Ex : « lorsque je n'ai pas les idées claires par rapport à un problème, l'une des premières choses que je fais est d'analyser la situation et prendre en considération toutes les informations pertinentes » ; « quand je suis fâché avec quelqu'un, d'habitude j'essaie de me mettre à sa place pour un moment » ; « j'essaie toujours d'analyser toutes les facettes d'un problème », etc.

⁵⁷⁵ Ex : « je suis à l'aise avec tous les types de personnes » ; « parfois je ressens de la vraie compassion pour tout le monde » ; « si je vois des personnes en danger, j'essaie de les aider d'une façon ou d'une autre », etc.

nécessaire pour le développement de la compréhension profonde des dynamiques intra-personnelles et inter-personnelles qui caractérisent la dimension cognitive de la sagesse. En ce qui concerne la dimension affective, la capacité à transcender sa propre perspective limitée en se mettant à la place de l'autre facilite le développement de sentiments de sympathie et de compassion envers autrui. Les trois dimensions de la sagesse se soutiennent mutuellement.

La sagesse, telle qu'elle est définie et mesurée par le questionnaire 3D-WS, a été spécialement étudiée par Ardeli chez des personnes âgées, dans un cadre gérontologique : l'hypothèse de travail est qu'il s'agit d'un facteur décisif pour un vieillissement heureux⁵⁷⁶. La sagesse est corrélée positivement et de façon significative avec le bien-être, la maîtrise, la présence de buts dans la vie et une évaluation subjective du niveau de santé. Elle est corrélée de façon négative avec des symptômes dépressifs, la perception de problèmes économiques et la peur de la mort. La sagesse est relativement stable dans le temps (fiabilité test-retest, après dix mois), et elle n'est pas corrélée avec la richesse, le fait d'être marié ou à la retraite, le sexe, les origines ethniques et le fait d'être hautement désirable d'un point de vue social (*social desirability index*). En général, la sagesse a un effet plus fort sur la satisfaction par rapport à sa propre vie que des indicateurs « objectifs » de bien-être, tels que le niveau de santé physique, le niveau socio-économique, la situation financière, l'environnement physique et l'intégration sociale. Cependant, à la différence de la satisfaction par rapport à sa propre vie, la sagesse en elle-même n'est pas strictement associée à ces indicateurs objectifs (à l'exception de la santé physique). Cela signifie que, chez l'individu sage la satisfaction n'est pas produite uniquement par des circonstances positives : le sage est en mesure d'affronter des vicissitudes en gardant une attitude positive.

Tout cela suggère que la sagesse, ainsi définie, peut effectivement jouer un rôle décisif pour un vieillissement heureux. Cependant, il est évident que les compétences cognitives – qui désignent la sagesse en tant que savoir – ne sont pas suffisantes pour que les effets de la sagesse se manifestent. Il est possible d'avoir des compétences cognitives plutôt développées, ainsi que des bonnes compétences affectives, sans être sage. Tel est le cas typique des membres des professions d'assistance psychologique et de conseil (psychologues cliniques, coach, psychanalystes, assistants sociaux). Ces individus possèdent une grande quantité de savoir sur la condition humaine, les vicissitudes typiques de la vie et sur la façon de gérer des conditions existentielles difficiles : typiquement, ils éprouvent aussi de la sympathie et de la compassion pour leur clients ou les personnes qu'ils assistent. Cependant, d'après Ardeli, souvent ces bons connaisseurs de la vie ne sont pas eux-mêmes des modèles de comportement à suivre. Il s'agit du syndrome du « *helpless*

⁵⁷⁶Baltes et Ardeli partagent le même arrière-plan scientifique.

helper »⁵⁷⁷ : à cause d'un manque de travail réflexif sur soi-même suffisamment intense, ces professionnels de l'aide psychologique ont souvent les mêmes problèmes que leurs clients. À défaut d'une auto-analyse continuelle, des connaissances psychologiques profondes risquent d'être utilisées pour produire des rationalisations qui justifient et renforcent les biais psychologiques de départ : c'est un phénomène général, qui ne concerne pas uniquement la sagesse. La solution est une expérience de plusieurs années dans la gestion des pièges et des faiblesses de sa propre psychologie. D'après Ardel, le composant cognitif de la sagesse, c'est-à-dire la capacité de jugement pratique, est parfaitement dissociable de la cohérence dans le comportement. Sur ce point elle rejoint les analyses de Sternberg et Baltes. Toutefois, il faut rappeler que, même si Ardel avait trouvé une corrélation entre savoir et action, il s'agirait uniquement d'une tendance générale, valide pour l'individu moyen, car son approche hérite de toutes les limitations de la théorie des traits classique. Dans ce scénario (qui n'est pas réel), on aurait prouvé seulement que le savoir de la sagesse a une certaine tendance fiable à unifier la personnalité.

Conclusion

À ce stade de notre analyse, les données psychologiques qui étaient la dissociation entre la sagesse en tant que jugement et la sagesse en tant que personnalité harmonieuse se multiplient. Lorsqu'on parle du « savoir » de la sagesse, on ne se réfère pas à des connaissances totalement abstraites, qu'on pourrait acquérir uniquement par l'étude d'une série de livres et répéter lors d'un examen ou d'une conférence. Si cela était le cas, la démonstration serait excessivement facile, car, bien sûr, ce savoir superficiel n'est pas en mesure de restructurer de façon profonde une personnalité. Par « acquisition du savoir de la sagesse » on entend un processus qui, par définition, change de façon profonde au moins la perception et le jugement d'une personne. En d'autres termes, le sage (au sens du jugement) perçoit le monde de façon différente et il est en mesure de mobiliser son savoir pour donner des conseils de façon efficace. Il ne va pas simplement répéter la maxime, le principe ou le proverbe les plus pertinents dans une situation donnée : il va juger de la spécificité de la situation et donner un conseil approprié, personnalisé et surtout concrètement efficace. En ce sens, il ne fait pas de doute que l'acquisition de la sagesse en tant que savoir est un processus long et difficile.

Néanmoins, cela n'est pas suffisant pour être en mesure d'appliquer soi-même l'enseignement de la sagesse : à l'état actuel de nos connaissances, la psychologie empirique contemporaine semble le montrer. La stratégie traditionnelle des philosophes qui défendent l'union des deux sens de sagesse

⁵⁷⁷ W. Schmidbauer, *Die Hilfflosen Helfer* (Rowohlt, 1977).

est plus sophistiquée qu'elle ne paraît à plusieurs commentateurs : on accuse très fréquemment la théorie de la vertu d'intellectualisme. Mais nous avons bien montré que le sens de « savoir » dont les anciens parlent entraîne déjà par définition une transformation profonde de la personnalité. Il reste à voir si la transformation coïncide avec celle qui définit la sagesse en tant que personnalité harmonieuse.

De prime abord, il n'y a pas de façon simple et immédiate de démontrer, contre les philosophes, que perception/jugement d'un côté et action efficace de l'autre sont dissociables. Au contraire, la thèse des philosophes est vraisemblable : quoi de mieux que l'expérience en première personne, si on veut développer la perception et l'imagination nécessaires pour comprendre profondément et juger correctement ? Quelle autre force aurait-elle pu forger la capacité de jugement, autre que le processus difficile par lequel on essaie, on fait face à l'échec, on corrige ses attitudes, jusqu'à atteindre le succès ? Répondre à ces questions n'est pas facile : c'est ici que les données empiriques se révèlent essentielles. Malgré la complexité des changements psychologiques impliqués par le développement d'un jugement véritablement « sage », cela ne suffit pas pour assurer l'harmonisation complète de la personnalité et la cohérence dans l'action. En plus, d'après les psychologues, la fragmentation et la spécialisation de l'excellence semblent s'appliquer également à la sagesse. La sagesse, qui devait nous fournir une réponse au problème de la fragmentation du caractère, est probablement elle-même fragmentées en plusieurs domaines.

Rien ne semble empêcher « la fragmentation de l'expertise » de s'étendre aussi au domaine moral. En termes d'unification et de complémentarité, le courage et la générosité n'occupent pas une place privilégiée par rapport à la capacité à jouer du violon ou à jouer aux échecs. La théorie de la vertu est contrainte à accepter un modèle de spécialisation morale.

Remarque conclusive : quelle place pour l'excellence du caractère dans la philosophie normative du XXIème siècle ?

À la fin de ce long périple théorique, nous espérons avoir montré la nécessité d'un dialogue entre la théorie de la vertu et la psychologie expérimentale, en vue de la naissance d'une nouvelle science interdisciplinaire de la vertu. Un volume considérable de littérature expérimentale peut fournir aux philosophes la modélisation adéquate de la personnalité qui leur a fait défaut pendant des siècles. En particulier, nous avons montré que la typologie en cours de développement par les psychologues socio-cognitifs va dans la bonne direction. La psychologie de l'excellence et de l'expertise peut contribuer à la mise au point de techniques pour le développement du caractère. Quant aux philosophes, ils peuvent offrir aux psychologues leurs compétences dans le débat normatif, après avoir mis en évidence le fait que la théorie de la vertu classique n'est pas une théorie de la personnalité aussi dépassée que la notion de phlogistique (à en juger des manuels de psychologie de la personnalité, c'est bien la conception que les psychologues se font de l'héritage de la philosophie !). La tradition philosophique est porteuse d'ambitions que nous sommes aujourd'hui en mesure d'évaluer pour la première fois.

Sans soutien mutuel, aussi bien la psychologie que la philosophie se destinent à un défaitisme injustifié et surtout prématuré. La psychologie empirique n'a pas le mandat (ni théorique, ni social) pour avancer un idéal substantiel, précis et ambitieux de caractère. La seule théorie empirique des vertus existant à l'heure actuelle témoigne très bien de cette tendance défaitiste⁵⁷⁸. La preuve du manque d'ambition est la suivante : dans ce modèle, les vertus sont conceptualisées à partir d'une notion de trait comme tendance générale et peu fiable à l'action. Cela n'a rien à voir avec la fiabilité de la performance excellente recherchée dans le cadre de la réflexion traditionnelle sur la vertu.

En ce qui concerne la philosophie, l'absence d'une modélisation rigoureuse du caractère joue à l'avantage des théories neutres et peu ambitieuses en termes psychologiques. Les difficultés rencontrées pour donner une assise solide à de déclarations théoriques retentissantes a fini par engendrer un certain embarras ambiant dans les rangs des représentants contemporains de l'éthique de la vertu : cela favorise sans doute les désertions du camp de l'excellence morale aristotélicienne vers le camp qui préconise la simple « décence du caractère »⁵⁷⁹.

⁵⁷⁸ Christopher Peterson et Martin E. P. Seligman, *Character strengths and virtues: a handbook and classification* (Washington, DC New York: American Psychological Association ; Oxford University Press, 2004).

⁵⁷⁹ Merritt, "Virtue Ethics and Situationist Personality Psychology."

La contribution de cette thèse a été avant tout un œuvre de déblayage conceptuel préparatoire, mais, dans la troisième partie, nous avons développé un premier essai de science interdisciplinaire de la vertu, en montrant que les ambitions du modèle aristotélicien classique ne sont pas tenables dans leur version originale. Il paraît très difficile de développer un caractère totalement harmonieux et il faut préférer une psychologie de la spécialisation morale : nous pouvons, au plus, bâtir notre caractère sur un nombre restreint de vertu, tout comme l'excellence pratique en général n'est possible que dans un nombre restreint de domaines.

Même si l'on reste dans le cadre de la psychologie morale, il s'agit avant tout d'un résultat négatif : nous avons surtout rejeté le modèle globaliste. Développer de manière substantielle et détaillée un modèle de spécialisation des vertus sera l'un des prochains objectifs de la science de la vertu naissante : cela demandera un grand travail empirique et conceptuel. En ce qui concerne l'éthique normative, la situation ne peut qu'être encore plus incertaine. Quelles sont les conséquences normatives de l'adoption d'un modèle de vertu en tant que spécialisation morale ? Comme nous l'avons déjà dit à plusieurs reprises, la marge de manœuvre théorique est considérable et il faudrait une autre thèse pour évaluer l'impact de cette psychologie morale en éthique normative. Néanmoins, dans cette conclusion, nous ne pouvons plus frustrer le lecteur davantage et il faut accorder une place à la spéculation. Sans scénarios de référence, l'esprit humain se fatigue à saisir et comprendre une thèse abstraite. Or, il vaut mieux qu'un modèle psychologique soit associé à des scénarios de référence spéculatifs qu'il ne reste mal compris.

Malgré les limitations dues à la nécessité d'une spécialisation, la psychologie morale des vertus que nous proposons est bien une tentative de relancer sur des bases solides un idéal d'excellence morale et non pas uniquement de décence morale.

Mais alors, quelle place pour l'excellence du caractère dans la philosophie normative du XXI^{ème} siècle ? Tout au long du présent travail, nous avons abordé les problèmes techniques, mais il semble que le climat intellectuel des sociétés démocratiques contemporaines soit hostile à l'idéal d'excellence du caractère surtout pour d'autres raisons. Ce serait, il paraît, un idéal incompatible avec le pluralisme et la liberté individuelle. De quoi pourrait-il s'agir, sinon d'une tentative d'imposition qui va limiter les choix des individus et des communautés ? L'histoire des propositions substantielles pour définir le fonctionnement psychophysique normal ou optimal de l'individu humain est, paraît-il, sombre. Aujourd'hui, la tentative de démontrer la supériorité de certaines valeurs pour l'individu et pour la société paraît de plus en plus arrogante et désespérée.

Pour que l'interaction et le dialogue soient possibles, les « milles voies au bonheur » qui constituent le pluralisme contemporain trouvent un terrain d'entente autour de la notion de « décence morale ».

La poursuite d'une voie idiosyncrasique à l'épanouissement doit être compatible avec la manifestation d'un certain nombre de tendances morales, le socle de base qui définit les attentes de la société vers le citoyen moyen. Mais cette « décence morale » est constamment revue à la baisse. Alors que le contenu substantiel de tout modèle d'optimalité s'érode au fur et à mesure que des façons différentes, créatives et intéressantes d'interpréter l'amour, l'amitié, l'identité culturelle, les relations interpersonnelles (etc.) sont mises au jour, il ne reste plus de place que pour une éthique minimale. Le développement d'un niveau décent de vertu morale, poursuivi par un grand nombre de théoriciens contemporains de la vertu⁵⁸⁰, ne semble pas représenter un but stable. Pour ne pas exclure les interprétations minoritaires de l'épanouissement individuel et collectif, la décence morale est vite vidée de tout contenu substantiel, et la vertu finit par disparaître complètement de la scène.

Le souci du respect du pluralisme et de la liberté individuelle nous contraint-il à abandonner toute notion de vertu (certainement en tant qu'excellence, mais après aussi, petit à petit, en tant que manifestation d'une moralité « décente »)? Ce serait ironique. Pour peu que la notion d'amélioration concernée soit cohérente au niveau conceptuel et justifiable au niveau pratique, il faut reconnaître que jamais nous n'avons eu autant de connaissances et de moyens pour « améliorer » l'être humain. La science nous détaille le fonctionnement du corps et de l'esprit et étudie et modélise les interactions sociales : l'histoire, l'anthropologie et la sociologie nous offrent un nombre illimité d'expériences naturelles et de cas d'études qui sont autant de leçons pour tester des hypothèses et prendre la mesure du potentiel humain. Nous allons posséder la science dont les défenseurs positivistes du projet de construction d'un Eden sur Terre pour l'homme (*paradise engineering*) auraient rêvé, en même temps que toute notion d'excellence objective est écartée de la scène intellectuelle ? Y a-t-il une sortie possible de cette impasse ? Pouvons-nous sauvegarder le pluralisme, la liberté individuelle et en même temps mettre en avant un idéal d'excellence morale ? Une solution possible réside dans la notion de spécialisation de l'excellence. Dans le respect du pluralisme et des inclinations individuelles, une éthique de l'excellence peut proposer plusieurs pistes de développement du caractère, chacune basée sur le développement d'un nombre limité de vertus. Si la stratégie marche, chaque acteur au sein de la société devrait pouvoir retrouver sa vocation parmi les modèles proposés.

Comment éviter que cette éthique de l'excellence plurielle ne finisse par avaliser et rationaliser *ex post* tout équilibre atteint par une communauté ou un individu ? Dans ce cas de figure, le modèle

⁵⁸⁰ Par exemple, Nussbaum, voir M. C. Nussbaum, *Women and human development: The capabilities approach* (Cambridge University Press, 2000).

d'excellence fabriqué sur mesure ne serait autre chose qu'une idéologie qui justifie le *statu quo*. L'éthique de l'excellence serait complètement vidée de sa force normative et critique.

C'est ici que le fondement empirique d'une science de la vertu donne sa contribution décisive. Grâce à la nature substantielle, informative et empirique de chacun des modèles de fonctionnement optimal proposés, il sera possible d'éviter une instrumentalisation idéologique. L'hypothèse selon laquelle un équilibre psychologique donné correspond à un mode excellent de fonctionnement de l'être humain aurait été testée et devra être cohérente avec l'ensemble de nos connaissances de psychologie positive. Les modèles d'excellence proposés correspondent à autant de façons de développer le potentiel psychologique humain pour atteindre un mode de fonctionnement supérieur. L'instrumentalisation pour des fins de justification idéologique sera difficile. En plus, l'excellence ambitieuse dont il est question n'existant pas à l'état naturel, aucun modèle de vertu spécialisée ne sera complaisant envers les individus qui s'y reconnaissent. Un entraînement pointu et un processus de changement et d'éducation difficiles seront nécessaires pour tout le monde.

Seul le développement substantiel de la science interdisciplinaire de la vertu que le présent travail préconise pourra juger de la plausibilité de cette stratégie. Il en va du sort d'un idéal pluriel d'excellence, le seul idéal d'excellence qui a un avenir au XXI^{ème} siècle.

Une éthique de l'excellence plurielle ne se limite pas à être compatible avec le pluralisme - tout en gardant sa force normative. Elle est en mesure d'approfondir la signification et l'impact du pluralisme des sociétés démocratiques contemporaines. Car une éthique de l'excellence plurielle doit nécessairement être assortie d'une théorie du « partage du travail moral » qui intensifie la cohésion et l'interdépendance des voix et des composants de la société. Tout comme la spécialisation de l'expertise dans les autres domaines nous force à une interdépendance accrue, l'approfondissement de la pluralité morale ira de pair avec une interdépendance plus radicale des différentes forces actives dans la société. Alors que la quête de la « décence » favorise un pluralisme de l'isolement (avec quelques passerelles, de plus en plus minimales et exiguës), la quête commune de l'excellence (par plusieurs voies) encourage un pluralisme de la cohésion et de l'interdépendance.

Une dernière remarque s'impose : la quête commune de l'excellence favorise le pluralisme dans un deuxième sens. Elle favorise la création de plus de variété et l'exploration de davantage d'options. Cela pourrait paraître paradoxal pour la raison suivante : même si l'excellence est plurielle, tout modèle psychologique informatif d'excellence nous pousse par définition à abandonner les modes de fonctionnement non excellents. De prime abord, cela réduit les options. Mais, en réalité, il faut concevoir la poursuite de l'excellence sur le mode de la montée d'un objet à l'intérieur d'un cône

renversé. Supposons que, à un moment t , un objet à l'intérieur de ce cône ne puisse bouger que dans le cercle obtenu par la section d'un plan parallèle à la base du cône. Poursuivre l'excellence correspond à monter dans le cône renversé (donc à s'approcher de la base). On abandonne la possibilité d'occuper tous les points de la section circulaire de départ, mais on se retrouve dans une autre section circulaire dont la superficie est plus grande. L'excellence exclut toutes les vieilles options de fonctionnement non optimal, mais elle en ouvre beaucoup plus de nouvelles. Car un modèle psychologique de spécialisation morale est insuffisant pour déterminer tous les aspects d'une culture : un grand nombre d'options peuvent s'articuler à partir du même modèle. Par contre, toute forme d'excellence est associée à certains fondamentaux psychologiques liés au fonctionnement optimal, créatif et productif de l'esprit humain. Une quête collective de l'excellence préconise une explosion de variation qui soutiendra une exploration de plus en plus poussée de l'espace du possible.

INDEX

- 3D-WS modèle, 423, 424
activation chronique, 129, 223
adaptés, 5, 159, 170, 207, 209, 216, 217
anti-réalisme, 97, 119, 168, 409
Approche typologique, 5, 199
approches révisionnistes, 4, 31
Approches socio-cognitives, 5, 132, 144, 160,
218, 222, 224, 225, 227, 229, 230, 232,
235, 236, 237, 242, 249, 378, 427
Ariston, 24, 399
Aristote, 8, 12, 14, 18, 22, 25, 26, 27, 28, 29,
30, 31, 34, 54, 55, 115, 148, 164, 200, 203,
265, 329, 371, 397, 398, 400, 401, 404,
405, 421
athlète moral, 149
attribution, 6, 7, 19, 33, 48, 51, 52, 64, 86, 87,
95, 96, 98, 102, 103, 104, 106, 107, 109,
111, 117, 118, 122, 127, 129, 147, 199,
257, 262, 310, 322, 323, 327, 341, 348,
349, 350, 351, 352, 397
automaticité pré-consciente, 377, 378, 381,
382, 387
biais de confirmation, 72, 107, 349, 352, 356,
358
biais psychologiques, 73, 425
bienveillance, 397, 398, 410
big five, 123, 138, 140, 161, 170, 171, 172,
173, 178, 183, 186, 189, 190, 191, 193,
195, 198, 202, 205, 212, 213, 237, 242,
252, 255, 415, 423
borderline, 6, 242, 243, 244
CAPS, 5, 232
caractère, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15,
16, 17, 18, 19, 20, 22, 23, 30, 33, 36, 37,
40, 41, 43, 44, 45, 47, 48, 49, 50, 51, 52,
53, 54, 55, 56, 58, 59, 62, 65, 68, 69, 84,
86, 89, 92, 96, 97, 98, 100, 101, 102, 106,
110, 122, 126, 131, 132, 135, 144, 147,
148, 149, 150, 151, 152, 155, 157, 161,
163, 164, 165, 166, 167, 168, 172, 174,
175, 176, 177, 179, 186, 188, 189, 190,
191, 193, 195, 196, 199, 200, 202, 203,
205, 209, 214, 215, 216, 217, 235, 237,
252, 254, 255, 256, 257, 258, 259, 263,
266, 268, 269, 270, 271, 273, 289, 290,
291, 295, 301, 303, 304, 305, 307, 310,
311, 325, 327, 332, 336, 337, 339, 340,
341, 342, 343, 344, 345, 346, 347, 348,
349, 350, 351, 353, 356, 358, 388, 389,
390, 391, 392, 393, 395, 397, 402, 405,
426, 427, 428, 429
Chrysippe, 399, 400
chunking, 366, 367
classification VIA, 5, 162, 188, 192, 193, 194,
195, 196
complaisance, 172, 415
concepts naïfs, 87, 115, 119
connexionnisme, 225
continuité absolue, 213, 214, 215
continuité centrée sur la personne, 213, 214,
215
continuité différentielle, 213
continuité structurelle, 213
contrainte, 31, 69, 172, 202, 219, 235, 236,
273, 304, 323, 343, 413, 426
cookbook, 286, 287
courage, 8, 9, 10, 11, 14, 23, 24, 25, 26, 27,
30, 33, 34, 48, 50, 54, 55, 57, 58, 59, 114,
115, 127, 134, 136, 137, 143, 146, 147,
148, 150, 154, 161, 163, 164, 166, 176,
177, 187, 188, 189, 194, 195, 204, 205,
206, 215, 254, 255, 260, 262, 263, 264,
280, 283, 287, 294, 295, 300, 321, 331,
342, 358, 360, 379, 388, 389, 398, 399,
400, 401, 402, 406, 416, 421, 426
cultures occidentales individualistes, 104
cultures orientales, 98
défaitisme, 34, 427
dispositionalisme, 89, 92
Dissonance cognitive, 6, 320
DSM, 192, 206, 244
effets top-down, 100, 126, 128, 129, 130
émotions positives, 16, 188, 192
engagements dynamiques, 4, 80, 82
entity theory, 99, 126
Equivalence fonctionnelle entre situations,
264, 265, 276, 284, 285
espèce naturelle psychologique, 39, 104, 108,
110, 133, 154, 190, 420
éthique de la vertu, 10, 12, 13, 34, 45, 46, 50,
53, 54, 55, 173, 203, 224, 336, 366, 427
excellence, 4, 5, 7, 8, 16, 17, 23, 24, 27, 28,
29, 30, 31, 34, 52, 57, 60, 105, 147, 160,
192, 194, 254, 255, 258, 260, 268, 296,
342, 343, 345, 360, 361, 362, 363, 364,
365, 366, 367, 368, 372, 374, 376, 377,

378, 384, 385, 387, 388, 390, 398, 399,
 403, 412, 416, 417, 419, 422, 426, 427,
 428, 429, 430
 excellence pratique, 18, 23, 160, 254, 255,
 259, 260, 374, 377, 378, 388, 398, 403,
 417, 428
 expertise, 4, 6, 7, 10, 14, 17, 23, 24, 25, 35,
 57, 67, 68, 72, 75, 78, 79, 80, 147, 148,
 149, 167, 254, 258, 260, 303, 360, 361,
 362, 365, 371, 374, 377, 378, 385, 386,
 387, 388, 389, 391, 392, 401, 412, 413,
 414, 415, 416, 417, 418, 419, 426, 427, 430
 explications intentionnelles, 94, 95, 96, 118
 extraversion, 38, 137, 140, 142, 158, 161,
 162, 171, 172, 183, 184, 186, 187, 189,
 190, 193, 198, 202, 203, 205, 220, 227,
 236, 252, 272, 280, 423
 facette, 169, 193
 fiabilisme, 4, 62, 67, 68, 74, 75
 fictionnalisme, 119, 120
 fitness, 133, 134, 135, 137, 157, 186
 générosité, 8, 10, 11, 24, 25, 34, 48, 49, 50,
 52, 101, 115, 187, 254, 256, 262, 263, 301,
 356, 358, 421, 426
 gentilhomme, 28, 29, 30, 31, 398
 gestion de la réputation, 38, 96, 142, 144,
 147, 158, 162, 224, 226
 globaliste, 4, 14, 17, 18, 20, 22, 100, 389,
 402, 428
 identité personnelle, 5, 38, 143, 144, 152,
 153, 155, 156, 236
 incremental theory, 100, 126
 interférences pragmatiques, 64, 65, 74, 96, 97,
 98
 interprétativisme, 119
 intuitions, 4, 14, 26, 53, 56, 58, 62, 63, 65, 66,
 67, 68, 70, 71, 72, 73, 74, 75, 76, 78, 79,
 80, 81, 82, 83, 84, 85, 87, 283, 293
 intuitions réfléchies, 4, 67, 72, 74
 intuitions robustes, 63, 72
 jurisprudence implicite, 82, 83
 KAPA, 5, 229, 230
 les traits spécifiques, 143
 loyauté, 108, 137, 189
 Méta-trait, 61
 Mischel Walter, 3, 105, 109, 110, 111, 125,
 159, 218, 219, 220, 227, 230, 232, 233,
 235, 236, 237, 240, 242, 273, 285, 286,
 287, 288, 293, 296, 297, 301, 302, 306
 modèle néo-aristotélicien, 4, 23, 28, 31, 157,
 254, 255, 390
 narcissisme, 232, 247
 neurobiologie, 84, 151
 névrotisme, 38, 137, 138, 140, 142, 159, 171,
 172, 177, 198, 202, 216, 227, 236, 252,
 415, 423
 ontogenèse, 5, 116, 117, 145
 Ouverture à l'expérience, 171, 207, 211
paradise engineering, 429
 partage du travail moral, 430
 personnalité, 3, 5, 6, 7, 8, 9, 14, 15, 17, 19,
 20, 31, 36, 37, 38, 39, 40, 45, 51, 52, 53,
 58, 61, 62, 95, 98, 100, 101, 102, 104, 126,
 127, 128, 129, 131, 132, 133, 134, 135,
 136, 138, 139, 140, 141, 142, 144, 145,
 146, 147, 148, 149, 150, 151, 152, 153,
 154, 155, 157, 158, 159, 160, 161, 162,
 163, 164, 165, 166, 167, 169, 170, 172,
 173, 175, 176, 177, 178, 180, 182, 183,
 184, 186, 187, 189, 190, 191, 196, 197,
 198, 199, 200, 203, 204, 206, 209, 210,
 212, 213, 214, 215, 216, 217, 218, 219,
 220, 222, 224, 226, 227, 229, 230, 231,
 232, 233, 234, 235, 236, 237, 238, 239,
 240, 242, 243, 248, 249, 250, 251, 252,
 253, 255, 257, 269, 270, 273, 280, 281,
 283, 286, 287, 290, 293, 295, 297, 298,
 304, 305, 306, 311, 314, 315, 316, 323,
 340, 346, 348, 349, 356, 358, 387, 391,
 392, 393, 396, 397, 398, 409, 411, 415,
 419, 420, 421, 422, 425, 426, 427
 personnalité animale, 140, 158, 162
 Personnalité dépendante, 6, 248
 philosophie de la psychologie
 wittgensteinienne, 46, 48, 50, 51, 52
 philosophie du canapé, 65, 71, 75, 80, 81
 philosophie expérimentale, 4, 65, 66, 67, 71,
 74, 75, 81
 phronesis, 26, 31, 395
 phylogenèse, 38, 132, 142, 145, 154, 229,
 230, 240, 242
 plasticité, 145, 228, 229, 255, 259, 364, 367,
 368, 412
 Platon, 8, 29, 30, 31, 296, 368
 populisme, 62, 68, 74, 76, 80, 81
 prédiction, 87, 89, 91, 98, 102, 106, 107, 108,
 117, 122, 123, 124, 127, 129, 180, 184,
 233, 236, 237, 243, 247, 287, 291, 293,
 294, 302, 315, 333, 334, 335, 349, 350

priming, 98, 241
 principes naïfs, 87
 prophétie auto-réalisatrice, 107, 144, 146, 241, 249
 prudence, 26, 27, 31, 34, 194, 395, 401, 404
 psychologie de l'automatisation, 259, 385
 Psychologie de la compétence, 7, 362
 psychologie de la personnalité, 5, 6, 8, 14, 15, 18, 19, 20, 38, 40, 44, 45, 104, 108, 115, 131, 132, 138, 140, 142, 150, 151, 152, 153, 154, 155, 157, 158, 159, 161, 162, 165, 166, 167, 170, 173, 174, 176, 179, 180, 186, 187, 189, 190, 193, 194, 196, 197, 198, 199, 201, 202, 204, 207, 209, 212, 214, 216, 218, 219, 230, 233, 234, 237, 254, 255, 257, 259, 262, 263, 269, 271, 272, 273, 281, 282, 283, 286, 287, 290, 295, 296, 297, 298, 300, 303, 310, 315, 360, 407, 408, 422, 427
 psychologie morale, 4, 5, 9, 11, 12, 13, 21, 22, 23, 30, 33, 34, 53, 55, 59, 60, 80, 82, 87, 149, 151, 161, 163, 164, 175, 255, 268, 360, 405, 428
 psychologie positive, 1, 15, 16, 17, 20, 188, 273, 430
 psychologie sociale, 8, 14, 18, 44, 45, 51, 61, 102, 108, 159, 218, 226, 241, 254, 255, 259, 262, 308, 309, 318, 321, 322, 324, 325, 330, 335, 348, 360
 réalisme, 5, 26, 40, 119, 120, 121, 168, 177, 181, 234
 réductionnisme, 5, 37, 131, 132, 133, 145, 149
 sagesse, 6, 7, 8, 10, 11, 21, 24, 26, 27, 28, 34, 50, 56, 57, 59, 93, 161, 187, 195, 255, 260, 275, 299, 303, 366, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 401, 402, 403, 404, 406, 407, 408, 409, 410, 411, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426
 Sagesse pratique, 31, 395
 scepticisme absolu, 47, 48
 scepticisme comparatif, 99, 115, 121
 scepticisme relatif, 47, 48
 schémas relationnels, 108, 231, 241, 257
 science de la personnalité, 5, 19, 36, 37, 100, 101, 131, 133, 146, 151, 152, 158, 159, 162, 198, 201, 205, 215, 216
 science de la vertu, 4, 8, 13, 15, 16, 17, 18, 20, 428, 430
 sélection fluctuante, 5, 133, 134, 136, 139
 sens commun, 4, 5, 7, 8, 9, 10, 11, 13, 15, 17, 18, 19, 29, 33, 34, 36, 37, 38, 46, 48, 49, 50, 51, 53, 54, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 76, 79, 80, 81, 82, 83, 84, 85, 86, 88, 89, 90, 91, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 107, 115, 116, 118, 120, 126, 128, 130, 131, 151, 152, 155, 161, 163, 164, 166, 173, 177, 191, 197, 199, 201, 213, 214, 231, 257, 264, 279, 280, 281, 303, 308, 316, 318, 321, 322, 326, 334, 339, 348, 349, 352, 358, 389, 390, 391, 396, 399, 402, 406, 407, 408, 409, 410, 412, 415, 416, 423
 situationnisme, 6, 102, 107, 110, 141, 216, 256, 262, 308, 309, 310, 324, 325, 328, 332, 333, 334, 335, 336, 337, 338, 339, 348, 350
 Socrate, 24, 368, 399
 soi, 10, 27, 36, 38, 56, 74, 76, 80, 81, 89, 101, 111, 112, 117, 125, 127, 131, 143, 144, 145, 148, 149, 150, 152, 154, 156, 158, 159, 160, 162, 171, 205, 210, 218, 219, 220, 223, 226, 230, 231, 232, 241, 243, 244, 246, 247, 249, 271, 284, 294, 295, 296, 297, 298, 300, 302, 304, 343, 374, 378, 392, 393, 394, 403, 409, 421, 423, 425
 soi cognitif, 149, 158, 230, 231, 232
 Sous-contrôlés, 207, 208
 Stabilité temporelle, 127
 stéréotype, 112, 113, 114, 122, 146, 197, 199, 246, 301, 348, 349, 352, 356, 377, 378, 379, 381
stereotype threat, 127, 246
 Stoïciens, 22, 29, 30, 399
 Sur-contrôlés, 207, 208
 tempérament, 37, 38, 132, 140, 142, 143, 144, 145, 146, 166, 204, 209, 227, 229, 255
 termes de trait, 9, 11, 22, 32, 33, 37, 47, 102, 103, 104, 105, 106, 115, 151, 161, 170, 171, 189, 190, 191, 192, 287, 303, 304, 349, 390
 théorie de l'entité, 126, 127, 128, 129
 théorie de l'évolution, 3, 38, 132, 133, 135, 141
 théorie de la vertu, 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 17, 18, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 34, 37, 39, 41, 42, 43, 45, 46, 47, 50, 54, 55, 57, 59, 61, 115, 131, 133, 138, 148, 151, 157, 158, 160, 161, 162,

163, 164, 165, 173, 176, 187, 188, 196,
 197, 199, 203, 209, 214, 215, 216, 217,
 229, 235, 254, 255, 258, 259, 260, 262,
 267, 268, 269, 273, 290, 296, 297, 300,
 303, 305, 306, 322, 336, 337, 339, 340,
 343, 347, 359, 360, 361, 372, 374, 388,
 390, 397, 398, 399, 415, 420, 426, 427
 théorie incrémentale, 126, 127
 théorie naïve, 4, 19, 85, 86, 87, 88, 91, 95,
 115
 théorie naïve spontanée, 87
 théories, 141
 trade-off, 134, 136, 138
 trait, 4, 5, 9, 10, 11, 14, 15, 19, 32, 33, 37, 38,
 41, 49, 50, 51, 52, 54, 64, 94, 98, 100, 102,
 104, 105, 106, 107, 108, 109, 110, 113,
 114, 115, 116, 117, 121, 122, 123, 124,
 127, 129, 131, 133, 136, 139, 142, 152,
 158, 159, 161, 162, 166, 167, 168, 169,
 170, 171, 172, 173, 174, 178, 179, 180,
 181, 182, 183, 185, 186, 187, 190, 193,
 195, 197, 198, 199, 200, 202, 203, 204,
 205, 207, 213, 214, 215, 218, 220, 224,
 233, 237, 262, 263, 264, 265, 266, 267,
 268, 270, 271, 272, 274, 275, 277, 279,
 280, 281, 289, 291, 294, 295, 296, 297,
 298, 299, 300, 301, 302, 303, 304, 305,
 306, 307, 311, 332, 340, 341, 342, 345,
 346, 348, 351, 354, 356, 358, 360, 379,
 380, 381, 390, 395, 406, 408, 409, 422,
 423, 427
 trait behavioriste, 123
 traits contextuels, 38, 132, 142, 144, 146
 Type A, personnalité, 201, 209
 Type B, personnalité, 201, 209
 types, 6, 18, 36, 37, 79, 94, 102, 106, 107,
 112, 123, 128, 132, 135, 136, 145, 146,
 151, 158, 159, 162, 163, 167, 174, 179,
 195, 197, 198, 199, 200, 201, 202, 203,
 204, 205, 206, 207, 208, 209, 213, 215,
 216, 217, 225, 229, 232, 235, 236, 237,
 238, 239, 251, 256, 265, 266, 274, 286,
 299, 326, 364, 375, 400, 406, 420, 423
 typologie, 5, 112, 132, 158, 159, 197, 198,
 199, 200, 201, 202, 203, 204, 205, 206,
 207, 209, 215, 216, 217, 232, 236, 237,
 238, 240, 248, 427
 utilitarisme, 85
 valeurs, 5, 24, 36, 37, 56, 101, 131, 132, 149,
 150, 163, 178, 198, 203, 205, 206, 219,
 236, 240, 241, 274, 293, 301, 329, 348,
 379, 381, 414, 416, 418, 420, 421, 428
 vertu, 1, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15,
 17, 18, 20, 22, 23, 24, 25, 26, 27, 28, 29,
 30, 31, 32, 33, 34, 39, 41, 43, 44, 46, 47,
 48, 50, 51, 53, 54, 55, 56, 57, 58, 59, 60,
 61, 65, 66, 68, 77, 147, 148, 149, 151, 157,
 158, 160, 161, 162, 164, 165, 173, 174,
 175, 176, 187, 188, 189, 190, 191, 192,
 193, 194, 195, 196, 203, 209, 214, 224,
 235, 243, 254, 255, 258, 259, 260, 267,
 268, 269, 273, 275, 280, 296, 297, 303,
 305, 307, 336, 340, 341, 343, 344, 345,
 360, 361, 368, 371, 372, 377, 381, 388,
 389, 390, 392, 396, 397, 398, 399, 400,
 401, 402, 404, 405, 406, 411, 413, 417,
 421, 427, 428, 429, 430
 Zenon, 24, 399

Bibliographie

Social justice research 17, no. 2 (2004).

Achenbaum, W. A. et L. Orwoll. "Becoming Wise: A Psycho-Gerontological Interpretation of the Book of Job." *International Journal of Aging and Human Development* 32, no. 1 (1991): 21-39.

Ackerman, P. L. "New Developments in Understanding Skilled Performance." *Current Directions in Psychological Science* 16, no. 5 (2007): 235-239.

Adams, Robert Merrihew. *A Theory of Virtue : Excellence in Being for the Good*. Oxford ; New York: Clarendon ; Oxford University Press, 2006.

Agrawal, H. R., J. Gunderson, B. M. Holmes et K. Lyons-Ruth. "Attachment Studies with Borderline Patients: A Review." *Harvard review of psychiatry* 12, no. 2 (2004): 94.

Ainsworth, M. D. S. *Patterns of Attachment: A Psychological Study of the Strange Situation*: Lawrence Erlbaum Associates, 1978.

Allport, Gordon W. *Personality; a Psychological Interpretation*. New York,: H. Holt and Company, 1937.

Anderson, C. A., K. B. Anderson, N. Dorr, K. M. DeNeve et M. Flanagan. "Temperature and Aggression." in *Advances in Experimental Social Psychology, Vol 32*, 63-133. San Diego: Academic Press Inc, 2000.

Anderson, J. R. "Automaticity and the Act-Super (*) Theory." *American Journal of Psychology* 105, no. 2 (1992): 165-180.

Annas, Julia. "Comments on John Doris's 'Lack of Character'." *Philosophy and Phenomenological Research* 71, no. 3 (2005): 636-642.

———. "Virtue as a Skill." *International journal of philosophical studies* 3, no. 2 (1995): 227-243.

———. "Virtue Ethics and Social Psychology." *A Priori* 2 (2003): 20-34.

Anscombe, G. E. M. "Modern Moral Philosophy." *Philosophy* 33, no. 124 (1958): 1-19.

Aristote. *Ethique À Nicomaque*. Traduit par J. Tricot: J. Vrin, 1972.

———. *Ethique À Nicomaque*. Traduit par J. Tricot: J. Vrin, 2007.

Arkin, R. M. et S. Duval. "Focus of Attention and Causal Attributions of Actors and Observers." *Journal of Experimental Social Psychology* 11, no. 5 (1975): 427-438.

Asch, S. E. "Studies of Independence and Conformity .1. A Minority of One against a Unanimous Majority." *Psychological Monographs* 70, no. 9 (1956): 1-70.

- Asendorpf, J. B., P. Borkenau, F. Ostendorf et M. A. G. Van Aken. "Carving Personality Description at Its Joints: Confirmation of Three Replicable Personality Prototypes for Both Children and Adults." *European Journal of Personality* 15, no. 3 (2001): 169-198.
- Athanassoulis, Nafsika. "A Response to Harman: Virtue Ethics and Character Traits." *Proceedings of the Aristotelian Society* (2000): 215-221.
- Atkinson, J. W. et D. Birch. "A Dynamic Theory of Action." New York: Wiley, 1970.
- Austin, W. "Sex-Differences in Bystander Intervention in a Theft." *Journal of Personality and Social Psychology* 37, no. 11 (1979): 2110-2120.
- Ayduk, Ö, V. Zayas, G. Downey, A. B. Cole, Y. Shoda et W. Mischel. "Rejection Sensitivity and Executive Control: Joint Predictors of Borderline Personality Features." *Journal of Research in Personality* 42, no. 1 (2008): 151-168.
- Ayduk, O., G. Downey et M. Kim. "Rejection Sensitivity and Depressive Symptoms in Women." *Personality and Social Psychology Bulletin* 27, no. 7 (2001): 868.
- Ayduk, O., G. Downey, A. Testa, Y. Yen et Y. Shoda. "Does Rejection Elicit Hostility in Rejection Sensitive Women." *Social Cognition* 17, no. 2 (1999): 245-271.
- Ayduk, O., R. Mendoza-Denton, W. Mischel, G. Downey, P. K. Peake et M. Rodriguez. "Regulating the Interpersonal Self: Strategic Self-Regulation for Coping with Rejection Sensitivity." *Journal of Personality and Social Psychology* 79, no. 5 (2000): 776-792.
- Badhwar, N. K. "The Limited Unity of Virtue." *Noûs* 30, no. 3 (1996): 306-329.
- Bagge, C., A. Nickell, S. Stepp, C. Durrett, K. Jackson et T. J. Trull. "Borderline Personality Disorder Features Predict Negative Outcomes 2 Years Later." *Journal of Abnormal Psychology* 113, no. 2 (2004): 279-288.
- Baldwin, M. W. "Relational Schemas." *The Coherence of Personality: Social-Cognitive Bases of Consistency, Variability, and Organization* (1999).
- Baltes, P. B. "On the Incomplete Architecture of Human Ontogeny. Selection, Optimization, and Compensation as Foundation of Developmental Theory." *The American Psychologist* 52, no. 4 (1997): 366-380.
- . "Theoretical Propositions of Life-Span Developmental Psychology: On the Dynamics between Growth and Decline." *Developmental Psychology* 23, no. 5 (1987): 611-626.
- Baltes, P. B. et U. M. Staudinger. "The Search for a Psychology of Wisdom." *Current Directions in Psychological Science* 2, no. 3 (1993): 75-80.
- Baltes, Paul. "Wisdom as Orchestration of Mind and Virtue." non publié, 2004.
- Bandura, A. "Self-Efficacy: Toward a Unifying Theory of Behavior Change." *Psychological Review* 84, no. 2 (1977): 191-215.

- Bargh, J. A. "Auto-Motives: Preconscious Determinants of Social Interaction." in *Handbook of Motivation and Cognition: Foundations of Social Behavior*, édité par E. Higgins et R. Sorrentino. New York: The Guilford Press, 1990.
- . "Conditional Automaticity: Varieties of Automatic Influence in Social Perception and Cognition." in *Unintended Thought*, édité par J. A. Bargh et J. Uleman. New York: Guilford, 1989.
- Bargh, J. A., W. J. Lombardi et E. T. Higgins. "Automaticity of Chronically Accessible Constructs in Person X Situation Effects on Person Perception: It's Just a Matter of Time." *Journal of Personality and Social Psychology* 55, no. 4 (1988): 599-605.
- Baron, R. A. "The Sweet Smell Of ... Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls." *Personality and Social Psychology Bulletin* 23, no. 5 (1997): 498-503.
- Baron, R. A. et J. Thomley. "A Whiff of Reality - Positive Affect as a Potential Mediator of the Effects of Pleasant Fragrances on Task-Performance and Helping." *Environment and Behavior* 26, no. 6 (1994): 766-784.
- Baumeister, R. F. et D. M. Tice. "Metatraits." *Journal of Personality* 56, no. 3 (1988): 571-598.
- Bégorre-Bret, Cyrille. "Aristote: Les Vertus Intellectuelles Entre Moralité Et Vérité." in *Le Retour Des Vertus Intellectuelles*, édité par Thomas Bénatouïl et Michel Le Du: Les cahiers philosophiques de Strasbourg, 2006.
- Beier, M. E. et P. L. Ackerman. "Age, Ability, and the Role of Prior Knowledge on the Acquisition of New Domain Knowledge: Promising Results in a Real-World Learning Environment." *Psychology and aging* 20, no. 2 (2005): 341-355.
- Bem, D. J. et A. Allen. "On Predicting Some of People Some of Time - Search for Cross-Situational Consistencies in Behavior." *Psychological Review* 81, no. 6 (1974): 506-520.
- Bem, D. J. et D. C. Funder. "Predicting More of People More of Time - Assessing Personality of Situations." *Psychological Review* 85, no. 6 (1978): 485-501.
- Bilalic, M., P. McLeod et F. Gobet. "Inflexibility of Experts—Reality or Myth? Quantifying the Einstellung Effect in Chess Masters." *Cognitive Psychology* 56, no. 2 (2008): 73-102.
- Bishop, Michael A. et J. D. Trout. *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press, 2005.
- Blass, T. "Understanding Behavior in the Milgram Obedience Experiment - the Role of Personality, Situations, and Their Interactions." *Journal of Personality and Social Psychology* 60, no. 3 (1991): 398-413.
- Blevins, G. A. et T. Murphy. "Feeling Good and Helping - Further Phonebooth Findings." *Psychological Reports* 34, no. 1 (1974): 326-326.
- Block, J. et N. Haan. *Lives through Time*: Bancroft Books Berkeley, 1971.

- Bornstein, R. F., H. M. Ng, H. A. Gallagher, D. M. Kloss et N. G. Regier. "Contrasting Effects of Self-Schema Priming on Lexical Decisions and Interpersonal Stroop Task Performance: Evidence for a Cognitive/Interactionist Model of Interpersonal Dependency." *Journal of Personality* 73, no. 3 (2005): 731-762.
- Bornstein, R. F., J. M. Riggs, E. L. Hill et C. Calabrese. "Activity, Passivity, Self-Denigration, and Self-Promotion: Toward an Interactionist Model of Interpersonal Dependency." *Journal of Personality* 64, no. 3 (1996).
- Borsboom, D., G. J. Mellenbergh et J. van Heerden. "The Theoretical Status of Latent Variables." *Psychological Review* 110, no. 2 (2003): 203-219.
- Boyd, R. et P. J. Richerson. "Not by Genes Alone: How Culture Transformed Human Evolution." Chicago: University of Chicago Press, 2004.
- Britt, T. W. "Metatraits - Evidence Relevant to the Validity of the Construct and Its Implications." *Journal of Personality and Social Psychology* 65, no. 3 (1993): 554-562.
- Broadie, S. *Ethics with Aristotle*: Oxford University Press, USA, 1991.
- Buchanan, T., J. A. Johnson et L. R. Goldberg. "Implementing a Five-Factor Personality Inventory for Use on the Internet." *European Journal of Psychological Assessment* 21, no. 2 (2005): 115-127.
- Budd, M. *Wittgenstein's Philosophy of Psychology*. London: Routledge, 1989.
- Buss, A. H. *Self-Consciousness and Social Anxiety*: WH freeman San Francisco, 1980.
- Buss, D. M. "Evolutionary Personality Psychology." *Annual Review of Psychology* 42, no. 1 (1991): 459-491.
- Buss, D. M. et D. P. Schmitt. "Sexual Strategies Theory: A Contextual Evolutionary Analysis of Human Mating." *Psychological Review* 100 (1993): 204-232.
- Cade, W. H. "Alternative Male Strategies: Genetic Differences in Crickets." *Science* 212, no. 4494 (1981): 563.
- Cantor, N. et J. F. Kihlstrom. *Personality and Social Intelligence*: Prentice-Hall.
- Cantor, N. et W. Mischel. "Prototypes in Person Perception." *Advances in experimental social psychology* 12 (1979): 3-52.
- Carus, P. et A. C. MacIntyre. *Dependent Rational Animals: Why Human Beings Need the Virtues*: Open Court Publishing, 2001.
- Carver, C. S. et M. F. Scheier. "Control Processes and Self-Organization as Complementary Principles Underlying Behavior." *The Dynamic Perspective in Personality and Social Psychology* 6, no. 4 (2002): 304-315.
- Caspi, A. "Personality Development across the Life Course." *Handbook of child psychology* 3 (1998): 311-388.

- . “Personality in the Life Course.” *Journal of Personality and Social Psychology* 53, no. 6 (1987): 1203-1213.
- Caspi, A. et T. E. Moffitt. “When Do Individual Differences Matter? A Paradoxical Theory of Personality Coherence.” *Psychological Inquiry* 4, no. 4 (1993): 247-271.
- Caspi, A. et B. W. Roberts. “Personality Development across the Life Course: The Argument for Change and Continuity.” *Psychological Inquiry* 12, no. 2 (2001): 49-66.
- Cattell, R. B. *Description and Measurement of Personality*: World book company, 1946.
- . “The Description of Personality: Basic Traits Resolved into Clusters.” *Journal of Abnormal and Social Psychology* 38, no. 4 (1943): 476-506.
- . *Personality and Motivation Structure and Measurement*: World Book, 1957.
- Cervone, D. “The Architecture of Personality.” *Psychological Review* 111, no. 1 (2004): 183-20421.
- . “Personality Architecture: Within-Person Structures and Processes.” *Annual Review of Psychology* 56, no. 1 (2005): 423-452.
- Chaiken, S. et Y. Trope. *Dual-Process Theories in Social Psychology*: Guilford Press New York, 1999.
- Chaplin, W. F. et L. R. Goldberg. “A Failure to Replicate the Bem and Allen Study of Individual-Differences in Cross-Situational Consistency.” *Journal of Personality and Social Psychology* 47, no. 5 (1984): 1074-1090.
- Chase, H. et H. A. Simon. “Skill in Chess.” *American Scientist* 61, no. 4 (1973): 394-403.
- Cheek, J. M., & Buss, A. H. “Shyness and Sociability.” *Journal of Personality and Social Psychology* 41 (1981): 330-339.
- Cheek, J. M. et E. N. Krasnoperova. “Varieties of Shyness in Adolescence and Adulthood.” in *Extreme Fear, Shyness, and Social Phobia: Origins, Biological Mechanisms, and Clinical Outcomes*, édité par L.A. Schmidt et J. Schulkin, 224–250. New York: Oxford University Press, 1999.
- Chen, C., M. Burton, E. Greenberger et J. Dmitrieva. “Population Migration and the Variation of Dopamine D4 Receptor (Drd4) Allele Frequencies around the Globe—a Family-Based Approach.” *Evolution and Human Behavior* 20, no. 5 (1999): 309-324.
- Choi, Incheol et Richard E. Nisbett. “Situational Salience and Cultural Differences in the Correspondence Bias and Actor-Observer Bias.” *Personality and Social Psychology Bulletin* 24, no. 9 (1998): 949-960.
- Choudhry, N. K., R. H. Fletcher et S. B. Soumerai. “Systematic Review: The Relationship between Clinical Experience and Quality of Health Care.” *Annals of Internal Medicine* 142, no. 4 (2005): 260.

- Church, A. T. "Culture and Personality: Toward an Integrated Cultural Trait Psychology." *Journal of Personality* 68, no. 4 (2000): 651-703.
- Clark, R. D. et L. E. Word. "Where Is Apathetic Bystander - Situational Characteristics of Emergency." *Journal of Personality and Social Psychology* 29, no. 3 (1974): 279-287.
- Clausen, J. A. et C. J. Jones. "Predicting Personality Stability across the Life Span: The Role of Competence and Work and Family Commitments." *Journal of Adult Development* 5, no. 2 (1998): 73-83.
- Clausen, J. S. "Adolescent Competence and the Shaping of the Life Course." *The American Journal of Sociology* 96, no. 4 (1991): 805-842.
- Clayton, V. "Erikson's Theory of Human Development as It Applies to the Aged: Wisdom as a Contradictive Cognition." *Human Development* 18 (1975): 119-128.
- Cloninger, C. R. "8 Completing the Psychobiological Architecture of Human Personality Development: Temperament, Character, and Coherence." *Understanding Human Development: Dialogues with Lifespan Psychology* (2003).
- Cooper, J. M. "The Unity of Virtue." *Social Philosophy and Policy* 15 (1998): 233-274.
- Coyne, J. C. et V. E. Whiffen. "Issues in Personality as Diathesis for Depression: The Case of Sociotropy-Dependency and Autonomy-Self-Criticism." *Psychological Bulletin* 118, no. 3 (1995): 358-378.
- Cunningham, Craig A. "A Certain and Reasoned Art." in *Character Psychology and Character Education*, édité par Daniel K. Lapsley et F. Clark Power. Notre Dame, Indiana: University of Notre Dame Press, 2005.
- Cunningham, M. R. "Weather, Mood, and Helping-Behavior - Quasi Experiments with the Sunshine Samaritan." *Journal of Personality and Social Psychology* 37, no. 11 (1979): 1947-1956.
- Dancy, Jonathan. *Ethics without Principles*. Oxford ; New York: Clarendon, 2004.
- Dar-Nimrod, Ilan et Steven J. Heine. "Exposure to Scientific Theories Affects Women's Math Performance." *Science* 314, no. 5798 (2006): 435-.
- Darley, J. M. et C. D. Batson. "From Jerusalem to Jericho - Study of Situational and Dispositional Variables in Helping Behavior." *Journal of Personality and Social Psychology* 27, no. 1 (1973): 100-108.
- David, P. et S. Samadi. *La Théorie De L'évolution*: Paris: Champus Université Flammarion, 2000.
- Dent, N. J. H. *The Moral Psychology of the Virtues*: Cambridge University Press, 1984.
- Deschavanne, E. et P. H. Tavoillot. "Philosophie Des Âges De La Vie." *Paris, Grasset, 2007*.
- Dingemanse, N. J., C. Both, P. J. Drent et J. M. Tinbergen. "Fitness Consequences of Avian Personalities in a Fluctuating Environment." *Proceedings: Biological Sciences* 271, no. 1541 (2004): 847-852.

- Doris, John M. *Lack of Character: Personality and Moral Behavior*. Cambridge, U.K. ; New York: Cambridge University Press, 2002.
- Downey, G. et O. Ayduk. "Correlates of Rejection Sensitivity in College Students." Données non publiées, 2002.
- Downey, G., S. Feldman et O. Ayduk. "Rejection Sensitivity and Male Violence in Romantic Relationships." *Personal Relationships* 7, no. 1 (2000): 45-61.
- Downey, G., A. L. Freitas, B. Michaelis et H. Khouri. "The Self-Fulfilling Prophecy in Close Relationships: Rejection Sensitivity and Rejection by Romantic Partners." *Journal of personality and social psychology* 75, no. 2 (1998): 545-560.
- Doyle, C. L. "Interpersonal Dynamics in Role Playing." *American Psychologist* 30 (1975): 1011-1013.
- Dreyfus, H. L. *On the Internet*: Routledge, 2001.
- Dreyfus, H. L. et S. E. Dreyfus. "The Ethical Implications of the Five-Stage Skill-Acquisition Model." *Bulletin of Science, Technology & Society* 24, no. 3 (2004): 251.
- Dreyfus, S. E. "The Five-Stage Model of Adult Skill Acquisition." *Bulletin of Science, Technology & Society* 24, no. 3 (2004): 177.
- Driver, Julia. *Uneasy Virtue, Cambridge Studies in Philosophy*. Cambridge, UK ; New York: Cambridge University Press, 2001.
- Dudycha, G. J. "An Objective Study of Punctuality in Relation to Personality and Achievement." *Archives of Psychology* 29 (1936): 1-53.
- Dwight, S. A., P. P. Wolf et J. H. Golden. "Metatraits: Enhancing Criterion-Related Validity through the Assessment of Traitedness." *Journal of Applied Social Psychology* 32, no. 10 (2002): 2202-2212.
- Eisenstadt, M. et Y. Kareev. "Aspects of Human Problem Solving: The Use of Internal Representations." *Explorations in cognition* (1975): 308-346.
- Elder Jr, G. H. "The Life Course as Developmental Theory." *Child Development* 69, no. 1 (1998): 1-12.
- Emmons, R. A. "Motives and Life Goals." *Handbook of personality psychology* (1997): 485-512.
- Ericsson, K. A. "Deliberate Practice and the Modifiability of Body and Mind: Toward a Science of the Structure and Acquisition of Expert and Elite Performance." *International Journal of Sport Psychology* 38 (2007): 4-34.
- . "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance." in *The Cambridge Handbook of Expertise and Expert Performance*, 683-703. New York, NY: Cambridge UP, 2006.
- Ericsson, K. A. et W. Kintsch. "Long-Term Working Memory." *Psychological Review* 102, no. 2 (1995): 211-245.

- Ericsson, K. A. et P. Ward. "Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance." *Current Directions in Psychological Science* 16, no. 6 (2007): 346-350.
- Erikson, E. *Childhood and Society. Revised and Enlarged*. New York: Norton, 1963.
- Eysenck, H. J. *A Model for Personality*: Springer, 1981.
- . *The Structure of Human Personality*: Methuen, 1970.
- Festinger, L. et J. M. Carlsmith. "Cognitive Consequences of Forced Compliance." *Journal of Abnormal and Social Psychology* 58, no. 2 (1959): 203-210.
- Field, C. A. et G. O'Keefe. "Behavioral and Psychological Risk Factors for Traumatic Injury." *Journal of Emergency Medicine* 26, no. 1 (2004): 27-35.
- Fishbach, A., R. S. Friedman et W. Arie. "Kruglanski (2003), "Leading Us Not into Temptation: Momentary Allurements Elicit Overriding Goal Activation,"." *Journal of Personality and Social Psychology* 84, no. 2: 296–309.
- Flanagan, Owen J. *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge, Mass.: Harvard University Press, 1991.
- Foot, P. *Virtues and Vices and Other Essays in Moral Philosophy*: Clarendon Press.
- Foucault, M. *Le Souci De Soi, Histoire De La Sexualité* Vol. 3. Paris: Gallimard, 1984.
- Franken, R. E., K. J. Gibson et P. Mohan. "Sensation Seeking and Disclosure to Close and Casual Friends." *Personality and individual differences* 11, no. 8 (1990): 829-832.
- Fredrickson Barbara, L. *Positivity*. New York: Random House, 2009.
- Frensch, P. A. et R. J. Sternberg. "Skill-Related Differences in Game Playing." in *Complex Problem Solving: Principles and Mechanisms*, édité par P. A. Frensch et R. J. Sternberg, 343-381. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.
- Fried, R. et L. Berkowitz. "Music Hath Charms ... And Can Influence Helpfulness." *Journal of Applied Social Psychology* 9, no. 3 (1979): 199-208.
- Friedman, M. et R. H. Rosenman. *Type a Behavior and Your Heart*: Knopf, 1974.
- Funder, D. C. "Global Traits - a Neo-Allportian Approach to Personality." *Psychological Science* 2, no. 1 (1991): 31-39.
- . "On the Accuracy of Personality Judgment - a Realistic Approach." *Psychological Review* 102, no. 4 (1995): 652-670.
- . "Personality." *Annual Review of Psychology* 52 (2001): 197-221.
- . "Towards a Resolution of the Personality Triad: Persons, Situations, and Behaviors." *Journal of Research in Personality* 40, no. 1 (2006): 21-34.

- Funder, D. C. et C. R. Colvin. "Explorations in Behavioral Consistency - Properties of Persons, Situations, and Behaviors." *Journal of Personality and Social Psychology* 60, no. 5 (1991): 773-794.
- Furr, R. M. et D. C. Funder. "Situational Similarity and Behavioral Consistency: Subjective, Objective, Variable-Centered, and Person-Centered Approaches." *Journal of Research in Personality* 38, no. 5 (2004): 421-447.
- Geher, G. et G. Miller. *Mating Intelligence: Sex, Relationships, and the Mind's Reproductive System*: Lawrence Erlbaum, 2007.
- Gilbert, D. T. et P. S. Malone. "The Correspondence Bias." *Psychological Bulletin* 117, no. 1 (1995): 21-38.
- Gollwitzer, P. M., H. Heckhausen et B. Steller. "Deliberative and Implemental Mind-Sets: Cognitive Tuning toward Congruous Thoughts and Information." *Journal of Personality and Social Psychology* 59, no. 6 (1990): 1119-1127.
- Goodall, J. *The Chimpanzees of Gombe: Patterns of Behavior*: Belknap Press of Harvard University Press, 1986.
- Gosling, S. D. "From Mice to Men: What Can We Learn About Personality from Animal Research?" *Psychological Bulletin* 127, no. 1 (2001): 45-86.
- Gosling, S. D. et O. P. John. "Personality Dimensions in Nonhuman Animals: A Cross-Species Review." *Current Directions in Psychological Science* 8, no. 3 (1999): 69-75.
- Grant, B. R. *Evolutionary Dynamics of a Natural Population: The Large Cactus Finch of the Galápagos*: University of Chicago Press, 1989.
- Greene, J. et J. Haidt. "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6, no. 12 (2002): 517-523.
- Groot, Adde. *Thought and Choice in Chess*: Mouton, 1978.
- Haan, N., R. Millsap et E. Hartka. "As Time Goes By: Change and Stability in Personality over Fifty Years." *Psychology and Aging* 1, no. 3 (1986): 220-232.
- Haidt, J. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108, no. 4 (2001): 814-834.
- Hanson, Jon et David Yosifon. "The Situation: An Introduction to the Situational Character, Critical Realism, Power Economics, and Deep Capture." In *Santa Clara University School of Law Legal Studies Research Papers Series*, 2006.
- Harari, H., O. Harari et R. V. White. "The Reaction to Rape by American Male Bystanders." *Journal of Social Psychology* 125, no. 5 (1985): 653-658.
- Harman, Gilbert. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99 (1999): 315-331.

- Hartshorne, Hugh. *Studies in the Organization of Character, Studies in the Nature of Character ; V.* 3. New York,: Macmillan, 1930.
- Hartshorne, Hugh et Mark Arthur May. *Studies in Deceit, Studies in the Nature of Character ; 1.* New York: Macmillan, 1928.
- Hauser, M. D., F. Cushman, L. Young et R. K. X. Jin. "A Dissociation between Moral Judgments and Justifications." *Mind and Language* 22 (2007): 1-21.
- Hecht, H. et D. R. Proffitt. "The Price of Expertise: Effects of Experience on the Water-Level Task." *Psychological Science* 6, no. 2 (1995): 90-95.
- Hershberger, S. L., R. Plomin et N. L. Pedersen. "Traits and Metraits - Their Reliability, Stability, and Shared Genetic Influence." *Journal of Personality and Social Psychology* 69, no. 4 (1995): 673-685.
- Higgins, Kathleen Marie. "Negative Virtues: Zhuangzi's Wuwei." in *Virtue Ethics Old and New*, édité par Stephen M. Gardiner: Cornell University Press, 2005.
- Hinsley, D. A., J. R. Hayes et H. A. Simon. "From Words to Equations: Meaning and Representation in Algebra Word Problems." *Cognitive processes in comprehension* (1977): 89-106.
- Hodges, B. H. et A. L. Geyer. "A Nonconformist Account of the Asch Experiments: Values, Pragmatics, and Moral Dilemmas." *Personality and Social Psychology Review* 10, no. 1 (2006): 2-19.
- Horvath, S. et C. C. Morf. "Implicitly Activated Worthlessness in Narcissism." Données non publiées: University of Bern, 2006.
- Hurka, T. "Virtuous Act, Virtuous Dispositions." *Analysis* 66, no. 289 (2006): 69-76.
- Hursthouse, R. *On Virtue Ethics*: Oxford University Press, 2000.
- Hutchinson, D. S. "Doctrines of the Mean and the Debate Concerning Skills in Fourth-Century Medicine, Rhetoric and Ethics." *Apeiron* 21, no. 2 (1988): 17-52.
- Hutto, D. D. "The Limits of Spectatorial Folk Psychology." *Mind and Language* 19, no. 5 (2004): 548-573.
- Hutton, Eric L. "Character, Situationism, and Early Confucian Thought." *Philosophical Studies* 127 (2006): 37-58.
- i Freixanet, G.M. "Personality Profile of Subjects Engaged in High Physical Risk Sports." *Personality and Individual Differences* 12, no. 10 (1991): 1087-1093.
- Isen, A. M. et P. F. Levin. "Effect of Feeling Good on Helping - Cookies and Kindness." *Journal of Personality and Social Psychology* 21, no. 3 (1972): 384-&.
- Jones, E. E. et K. E. Davis. "From Acts to Dispositions - the Attribution Process in Person Perception." *Advances in Experimental Social Psychology* 2, no. 4 (1965): 219-266.

- Jones, Edward E. et Victor A. Harris. "The Attribution of Attitudes." *Journal of Experimental Social Psychology* 3, no. 1 (1967): 1-24.
- Joyce, R. *The Evolution of Morality*: MIT Press, 2006.
- Kagan, J. *In: Galen's Prophecy: Temperament in Human Nature* New York: BasicBooks, 1994.
- Kamtekar, Rachana. "Situationism and Virtue Ethics on the Content of Our Character." *Ethics: An International Journal of Social, Political, and Legal Philosophy* 114, no. 3 (2004): 458-491.
- . "Updating Practical Wisdom." in *Virtue Ethics, Character and Moral Psychology*, édité par C. Upton. Oxford: Oxford University Press, à paraître.
- Kauppinen, A. "The Rise and Fall of Experimental Philosophy." *Philosophical Explorations* 10, no. 2 (2007): 95-118.
- Kegan, R. *The Evolving Self: Problem and Process in Human Development*: Harvard University Press, 1982.
- Keirsey, D. et M. M. Bates. *Please Understand Me: Character & Temperament Types*: Prometheus Nemesis Book Co., 1984.
- Kenrick, D. T. et D. C. Funder. "Profiting from Controversy - Lessons from the Person-Situation Debate." *American Psychologist* 43, no. 1 (1988): 23-34.
- King, J. E. et A. J. Figueredo. "The Five-Factor Model Plus Dominance in Chimpanzee Personality." *Journal of Research in Personality* 31, no. 2 (1997): 257-271.
- Klein, G. "The Power of Intuition." *Currency-Doubleday, New York, NY* (2003).
- Klein, G. A. *Sources of Power: How People Make Decisions*: MIT Press, 1998.
- Klirs, E. G. et W. Revelle. "Predicting Variability from Perceived Situational Similarity." *Journal of Research in Personality* 20, no. 1 (1986): 34-50.
- Kohlberg, L. *The Psychology of Moral Development: The Nature and Validity of Moral Stages*: Harper & Row, 1984.
- Krahé, Barbara. *Situation Cognition and Coherence in Personality : An Individual-Centered Approach, European Monographs in Social Psychology*. Cambridge England ; New York Paris: Cambridge University Press ; Editions de la Maison des Sciences de l'Homme, 1990.
- Kuhl, J. et S. L. Koole. "Workings of the Will." *Handbook of Experimental Existential Psychology* (2004).
- Kurz-Milcke, E. et G. Gigerenzer. *Experts in Science and Society*: Springer, 2004.
- Lank, D. B., C. M. Smith, O. Hanotte, T. Burke et F. Cooke. "Genetic Polymorphism for Alternative Mating Behaviour in Lekking Male Ruff *Philomachus Pugnax*." *Nature* 378, no. 6552 (1995): 59-62.
- Latane, B. et J. Rodin. "A Lady in Distress - Inhibiting Effects of Friends and Strangers on Bystander Intervention." *Journal of Experimental Social Psychology* 5, no. 2 (1969): 189-&.

- Latané, Bibb et John M. Darley. *The Unresponsive Bystander: Why Doesn't He Help?*, Century Psychology Series. New York,: Appleton-Century Crofts, 1970.
- Lepock, Chris. "Metacognition and Epistemic Virtue." thèse de doctorat: University of Alberta, 2007.
- Levinson, D. *The Seasons of a Man's Life* New York: Alfred A. Knopf Inc, 1978.
- Levinson, D. J. "A Conception of Adult Development." *American Psychologist* 41, no. 1 (1986): 3-13.
- Lewin, Kurt, Donald Keith Adams et Karl Edward Zener. *A Dynamic Theory of Personality; Selected Papers*. 1st édition, *Mcgraw-Hill Publications in Psychology*. New York, London,: McGraw-Hill Book Company inc., 1935.
- Little, B. R. "Personal Projects: A Rationale and Method for Investigation." *Environment and Behavior* 15, no. 3 (1983): 273.
- Loevinger, J. "Measurement of Personality: True or False." *Psychological Inquiry* 4, no. 1 (1993): 1-16.
- . "Stages of Personality Development." *Handbook of personality psychology* (1997): 199–208.
- Loevinger, J. et A. Blasi. *Ego Development: Conceptions and Theories*: Jossey-Bass, 1976.
- Logan, G. D. "Toward an Instance Theory of Automatization." *Psychological Review* 95, no. 4 (1988): 492-527.
- Lord, C. G. "Predicting Behavioral Consistency from an Individuals Perception of Situational Similarities." *Journal of Personality and Social Psychology* 42, no. 6 (1982): 1076-1088.
- Lord, C. G., L. Ross et M. R. Lepper. "Biased Assimilation and Attitude Polarization - Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37, no. 11 (1979): 2098-2109.
- Lovibond, S. H., Mithiran et W. G. Adams. "Effects of 3 Experimental Prison Environments on the Behavior of Nonconvict Volunteer Subjects." *Australian Psychologist* 14, no. 3 (1979): 273-287.
- Luchins, A. S. "Mechanization in Problem Solving; the Effect of Einstellung." (1942).
- Lund, O. C. H., C. K. Tamnes, C. Moestue, D. M. Buss et M. Vollrath. "Tactics of Hierarchy Negotiation." *Journal of Research in Personality* 41, no. 1 (2007): 25-44.
- Lynch, M. et B. Walsh. *Genetics and Analysis of Quantitative Traits*, 1998.
- Machery, Edouard. "Concepts Are Not a Natural Kind." *Philosophy of Science*, no. 72 (2005): 444–467.
- Mackie, J. L. *Ethics: Inventing Right and Wrong*: Penguin, 1977.

- Macrae, C. N. et G. V. Bodenhausen. "Social Cognition: Thinking Categorically About Others." *Annual Review of Psychology* 51, no. 1 (2000): 93-120.
- Magnusson, D., M. Gerzen et B. Nyman. "Generality of Behavioral Data-I - Generalization from Observations on One Occasion." *Multivariate Behavioral Research* 3, no. 3 (1968): 295-320.
- Malle, B. F. "The Actor-Observer Asymmetry in Attribution: A (Surprising) Meta-Analysis." *Psychological Bulletin* 132, no. 6 (2006): 895-919.
- Malle, B. F., J. M. Knobe et S. E. Nelson. "Actor-Observer Asymmetries in Explanations of Behavior: New Answers to an Old Question." *Journal of Personality and Social Psychology* 93, no. 4 (2007): 491-514.
- Mameli, M. et P. Bateson. "Innateness and the Sciences." *Biology and Philosophy* 21, no. 2 (2006): 155-188.
- Mantell, D. M. "Potential for Violence in Germany." *Journal of Social Issues* 27, no. 4 (1971): 101-112.
- Marcus, Gary. *Kluge: The Haphazard Construction of the Human Mind*: Houghton Mifflin Co, 2008.
- Markus, H. et E. Wurf. "The Dynamic Self-Concept: A Social Psychological Perspective." *Annual Review of Psychology* 38, no. 1 (1987): 299-337.
- Mathews, K. E. et L. K. Canon. "Environmental Noise-Level as a Determinant of Helping Behavior." *Journal of Personality and Social Psychology* 32, no. 4 (1975): 571-577.
- McArthur, L. Z. et D. L. Post. "Figural Emphasis and Person Perception." *Journal of Experimental Social Psychology* 13, no. 6 (1977): 520-535.
- McCrae, R. R. et P. T. Costa Jr. "A Five-Factor Theory of Personality." *Handbook of personality: Theory and research* 2 (1999): 139-153.
- McCrae, R. R., P. T. Costa Jr, F. Ostendorf, A. Angleitner, M. Hrebickova, M. D. Avia, J. Sanz, M. L. Sanchez-Bernardos, M. E. Kusdil et R. Woodfield. "Nature over Nurture: Temperament, Personality, and Life Span Development." *Journal of Personality and Social Psychology* 78, no. 1 (2000): 173-186.
- McCrae, R. R. et P. T. Costa. "Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model." *The five-factor model of personality: Theoretical perspectives* (1996): 51-87.
- McDowell, J. "Virtue and Reason." *The Monist* 62, no. 3 (1979): 331-350.
- McKenzie, J. "Neuroticism and Academic Achievement: The Furneaux Factor." *Personality and individual differences* 10, no. 5 (1989): 509-515.
- McKenzie, J., M. Taghavi-Khonsary et G. Tindell. "Neuroticism and Academic Achievement: The Furneaux Factor as a Measure of Academic Rigour." *Personality and Individual Differences* 29, no. 1 (2000): 3-11.

- Medin, D. L. et S. Atran. *Folkbiology*: Bradford Book, 1999.
- Meehl, P. E. "Wanted—a Good Cookbook." *American Psychologist* 11 (1956): 263-272.
- Mendoza-Denton, R., O. N. Ayduk, Y. Shoda et W. Mischel. "Cognitive-Affective Processing System Analysis of Reactions to the Oj Simpson Criminal Trial Verdict." *Journal of Social Issues* 53, no. 3 (1997): 563-581.
- Mendoza-Denton, R., G. Downey, V. J. Purdie, A. Davis et J. Pietrzak. "Sensitivity to Status-Based Rejection: Implications for African American Students' College Experience." *Journal of Personality and Social Psychology* 83, no. 4 (2002): 896-918.
- Merritt, Maria. "Virtue Ethics and Situationist Personality Psychology." *Ethical Theory and Moral Practice* V3, no. 4 (2000): 365-383.
- Metcalf, J. et W. Mischel. "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower." *Psychological Review* 106, no. 1 (1999): 3-19.
- Mikhail, J. "Universal Moral Grammar: Theory, Evidence and the Future." *Trends in Cognitive Sciences* 11, no. 4 (2007): 143-152.
- Milgram, S. "Behavioral-Study of Obedience." *Journal of Abnormal Psychology* 67, no. 4 (1963): 371.
- Milgram, Stanley. *Obedience to Authority; an Experimental View*. 1ère édition. New York: Harper & Row, 1974.
- Miller, G. F. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*: Doubleday Books, 2000.
- Mischel, W. "Toward a Cognitive Social Learning Reconceptualization of Personality." *Psychology Review* 80, no. 4 (1973): 252-283.
- Mischel, W. et C. C. Morf. "The Self as a Psycho-Social Dynamic Processing System: A Meta-Perspective on a Century of the Self in Psychology." *Handbook of self and identity* (2003): 15-43.
- Mischel, W., A. R. Zeiss et E. B. Ebbesen. "Cognitive and Attentional Mechanism in Delay of Gratification." *Journal of Personality and Social Psychology* 21, no. 2 (1972): 204.
- Mischel, Walter. "The Interaction of Person and Situation." in *Personality at the Crossroads: Current Issues in Interactional Psychology*, édité par S. Endler D. Magnusson. Hillsdale, NJ: Erlbaum, 1977.
- . *Introduction to Personality : Toward an Integration*. 7th édition. Hoboken, NJ: J. Wiley & Sons, 2004.
- . *Personality and Assessment*. Mahwah, NJ: L. Erlbaum Associates, 1968.
- . "Toward an Integrative Science of the Person." *Annual Review* 55, no. 1 (2004): 1-22.

- Mischel, Walter et Philip K. Peake. "Beyond Deja-Vu in the Search for Cross-Situational Consistency." *Psychological Review* 89, no. 6 (1982): 730-755.
- Mischel, Walter et Yuichi Shoda. "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure." *Psychological Review* 102, no. 2 (1995): 246-268.
- Molden, Daniel C. et Carol S. Dweck. "Finding "Meaning" In Psychology - a Lay Theories Approach to Self-Regulation, Social Perception, and Social Development." *American Psychologist* 61, no. 3 (2006): 192-203.
- Monahan, Liam Murphy. "Meeting Anscombe's Demand: Toward a Moral Psychology of Character." University of Notre Dame, 2004.
- Morf, C. C., M. Davidov et D. Ansara. "Strategic Regulation of Desired Identity Images in Response to Opportunities and Threats: Effects of Audience Characteristics on Narcissistic Self-Presentation." *Manuscript under review* (2006).
- Morf, C. C. et F. Rhodewalt. "Narcissism and Self-Evaluation Maintenance: Explorations in Object Relations." *Personality and Social Psychology Bulletin* 19, no. 6 (1993): 668.
- Morf, C. C., C. Weir et M. Davidov. "Narcissism and Intrinsic Motivation: The Role of Goal Congruence." *Journal of Experimental Social Psychology* 36, no. 4 (2000): 424-438.
- Moriarty, T. "Crime, Commitment, and Responsive Bystander - 2 Field Experiments." *Journal of Personality and Social Psychology* 31, no. 2 (1975): 370-376.
- Morizot, Julien. "Le Développement De La Personnalité De L'homme De L'adolescence Au Milieu De La Vie : Approches Centrées Sur Les Variables Et Sur Les Personnes." Thèse de doctorat, 2003.
- Nahmias, E., S. Morris, T. Nadelhoffer et J. Turner. "Surveying Freedom: Folk Intuitions About Free Will and Moral Responsibility." *Philosophical Psychology* 18, no. 5 (2005): 561-584.
- Natali, C. *L'action Efficace: Études Sur La Philosophie De L'action D'aristote*: Peeters, 2004.
- . *La Sagesza Di Aristotele*: Bibliopolis, 1989.
- Nelkin, Dana K. "Freedom, Responsibility and the Challenge of Situationism." in *Free Will and Moral Responsibility*, 181-206. Malden: Blackwell 2005.
- Nelson, T. O. et L. Narens. "Metamemory: A Theoretical Framework and New Findings." *The psychology of learning and motivation* 26 (1990): 125-173.
- Nettle, D. "The Evolution of Personality Variation in Humans and Other Animals." *American Psychologist* 61, no. 6 (2006): 622-631.
- . "An Evolutionary Approach to the Extraversion Continuum." *Evolution and Human Behavior* 26, no. 4 (2005): 363-373.
- Nettle, D. et H. Clegg. "Schizotypy, Creativity and Mating Success in Humans." *Proceedings: Biological Sciences* 273, no. 1586 (2006): 611-615.

- Neugarten, B. L. "Personality and the Aging Process." *Gerontologist* 12, no. 1 (1972): 9-15.
- . "Time, Age, and the Life Cycle." *Gerontologist* 136, no. 7 (1979): 887-894.
- Newcomb, Theodore Mead. "The Consistency of Certain Extrovert-Introvert Behavior Patterns in 51 Problem Boys." Teachers College Columbia University, 1929.
- Nichols, S., S. Stich et J. Weinberg. "Metaskepticism: Meditations in Ethno-Epistemology." *The Skeptics* (2003): 227-247.
- Nussbaum, M. "Non-Relative Virtues: An Aristotelian Approach." (1993).
- Nussbaum, M. C. "Virtue Ethics: A Misleading Category?" *The Journal of Ethics* 3, no. 3 (1999): 163-201.
- . *Women and Human Development: The Capabilities Approach*. Cambridge University Press, 2000.
- O'Steen, S., A. J. Cullum et A. F. Bennett. "Rapid Evolution of Escape Performance in Trinidadian Guppies (*Poecilia Reticulata*)." *Evolution* 56 (2002): 776-784.
- Oakley, J. et D. Cocking. *Virtue Ethics and Professional Roles*. Cambridge University Press, 2001.
- Orlando, N. J. "The Mock Ward: A Study in Simulation." in *Behavior Disorders: Perspectives and Trends*, édité par O. Milton et R. G. Wahler. Philadelphia: Lippincott, 1973.
- Peake, P. K. "Searching for Consistency: The Carleton Student Behavior Study." Thèse de doctorat, 1982.
- Pervin, L. A. *Goal Concepts in Personality and Social Cognition*. Hillsdale, NJ: Erlbaum, 1989.
- . *The Science of Personality*. New York: Wiley, 1996.
- Peterson, C. *A Primer in Positive Psychology*. New York: Oxford University Press, 2006.
- Peterson, Christopher et Martin E. P. Seligman. *Character Strengths and Virtues: A Handbook and Classification*. Washington, DC New York: American Psychological Association ; Oxford University Press, 2004.
- Pietrzak, J., G. Downey et O. Ayduk. "Rejection Sensitivity as an Interpersonal Vulnerability." *Interpersonal cognition* (2005): 62-84.
- Piliavin, Jane Allyn. *Emergency Intervention*. New York: Academic Press, 1981.
- Pinel, E. C. "Stigma Consciousness: The Psychological Legacy of Social Stereotypes." *Journal of Personality and Social Psychology* 76, no. 1 (1999): 114-128.
- Prinz, J. J. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford; New York: Oxford University Press, 2004.
- Prinz, Jesse. *The Emotional Construction of Morals*. Oxford: Oxford University Press, 2007.

- Pritchard, D. H. "Recent Work on Radical Skepticism." *American Philosophical Quarterly* 39, no. 3 (2002): 215-257.
- Purdie, V. et G. Downey. "Rejection Sensitivity and Adolescent Girls' Vulnerability to Relationship-Centered Difficulties." *Child Maltreatment* 5, no. 4 (2000): 338.
- Quine, W. V. *Le Mot Et La Chose*. Traduit par J. Dopp et P Gochet: Flammarion Paris, 1999.
- Rawls, J. *Théorie De La Justice*. Paris Le Seuil, 1987.
- Reder, L. M. "Selection Strategies in Question Answering." *Cognitive Psychology* 19 (1987): 90-138.
- Reder, L. M. et C. D. Schunn. "Bringing Together the Psychometric and Strategy Worlds: Predicting Adaptivity in a Dynamic Task." *Cognitive regulation of performance: Interaction of theory and application. Attention and Performance XVII* (1999): 315–342.
- Reicher, S et S. A Haslam. "Rethinking the Psychology of Tyranny." 2004.
- Rhodewalt, F. et D. L. Sorrow. "Interpersonal Self-Regulation: Lessons from the Study of Narcissism." *Handbook of self and identity* (2003): 519–535.
- Roberts, B. W. et W. F. DelVecchio. "The Rank-Order Consistency of Personality Traits from Childhood to Old Age: A Quantitative Review of Longitudinal Studies." *Psychological Bulletin* 126, no. 1 (2000): 3-25.
- Robins, R. W., O. P. John et A. Caspi. "The Typological Approach to Studying Personality." *Methods and models for studying the individual* (1998): 135–160.
- Rogosch, F. A. et D. Cicchetti. "Child Maltreatment, Attention Networks, and Potential Precursors to Borderline Personality Disorder." *Development and Psychopathology* 17, no. 04 (2005): 1071-1089.
- Rosental, R. et D. B. Rubin. "Interpersonal Expectancy Effects: The First 345 Studies"." *Behavioral and Brain Science* 3 (1978): 377-415.
- Ross, L. D. "Getting Down to Fundamentals: Lay Dispositionism and the Attributions of Psychologists." *Psychological Inquiry* 12, no. 1 (2001): 37-40.
- Ross, L., M. R. Lepper et M. Hubbard. "Perseverance in Self-Perception and Social Perception - Biased Attributional Processes in Debriefing Paradigm." *Journal of Personality and Social Psychology* 32, no. 5 (1975): 880-892.
- Ross, Lee D., Teresa M. Amabile et Julia L. Steinmetz. "Social Roles, Social-Control, and Biases in Social-Perception Processes." *Journal of Personality and Social Psychology* 35, no. 7 (1977): 485-494.
- Ross, Lee et Richard E. Nisbett. *The Person and the Situation: Perspectives of Social Psychology, McGraw-Hill Series in Social Psychology*. New York: McGraw-Hill, 1991.
- Royce, J. R. "On the Construct Validity of Open-Field Measures." *Psychological Bulletin* 84, no. 6 (1977): 1098-1106.

- Rudolph, K. D., C. Hammen et D. Burge. "A Cognitive-Interpersonal Approach to Depressive Symptoms in Preadolescent Children." *Journal of Abnormal Child Psychology* 25, no. 1 (1997): 33-45.
- Rusten, J., I. C. Cunningham et A. D. Knox. "Theophrastus: Characters, Mimes, Cercidas, and the Choliambic Poets." Cambridge, MA: Harvard University Press, 1993.
- Sabini, J. et M. Silver. "Lack of Character? Situationism Critiqued." *Ethics* 115, no. 3 (2005): 535-562.
- Sabini, John, Michael Siepmann et Julia Stein. "The Really Fundamental Attribution Error in Social Psychological Research." *Psychological Inquiry* 12, no. 1 (2001): 1-15.
- Samuels, J., O. J. Bienvenu, B. Cullen, P. T. Costa, W. W. Eaton et G. Nestadt. "Personality Dimensions and Criminal Arrest." *Comprehensive Psychiatry* 45, no. 4 (2004): 275-280.
- Schwarz, L., K. Jennings, J. Petrillo et R. F. Kidd. "Role of Commitments in the Decision to Stop a Theft." *Journal of Social Psychology* 110, no. 2 (1980): 183-192.
- Selman, R. L. *The Growth of Interpersonal Understanding: Developmental and Clinical Analyses*: Academic Press, 1980.
- Shaffer, D. R., M. Rogel et C. Hendrick. "Intervention in Library - Effect of Increased Responsibility on Bystanders Willingness to Prevent a Theft." *Journal of Applied Social Psychology* 5, no. 4 (1975): 303-319.
- Shand, A. F. *The Foundations of Character: Being a Study of the Tendencies of the Emotions and Sentiments*: Macmillan, 1914.
- Sheldon, W. H. "The Varieties of Human Temperament." New York: Harper Row, 1942.
- Sherman, N. *Making a Necessity of Virtue*: Cambridge University Press New York, 1997.
- Shoda, Y., W. Mischel et J. C. Wright. "The Role of Situational Demands and Cognitive Competences in Behavior Organization and Personality Coherence." *Journal of Personality and Social Psychology* 65, no. 5 (1993): 1023-1035.
- Shoda, Yuichi, Scott LeeTiernan et Walter Mischel. "Personality as a Dynamical System: Emergence of Stability and Distinctiveness from Intra- and Interpersonal Interactions." *Personality & Social Psychology Review* 6, no. 4 (2002): 316-325.
- Snow, Nancy. "Habitual Virtuous Actions and Automaticity." *Ethical Theory and Moral Practice*.
- Snyder, M. "Self-Monitoring of Expressive Behavior." *Journal of Personality and Social Psychology* 30, no. 4 (1974): 526-537.
- Snyder, M. et W. B. Swann. "Behavioral Confirmation in Social Interaction: From Social Perception to Social Reality." *Journal of Experimental Social Psychology* 14 (1978): 148-162.
- Snyder, Mark et Olivier Klein. "Construing and Constructing Others: On the Reality and the Generality of the Behavioral Confirmation Scenario." *Interaction Studies* 6 (2005): 53-67.

- Sosa, E. "Experimental Philosophy and Philosophical Intuition." *Philosophical Studies* 132, no. 1 (2007): 99-107.
- . *Knowledge in Perspective: Selected Essays in Epistemology*: Cambridge University Press, 1991.
- Sowarka, D. "Weisheit Und Weise Personen: Common-Sense-Konzepte Alterer Menschen [Wisdom and Wise Persons: Common Sense Concepts in Older People]." *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 21 (1989): 87-109.
- Sowarka, D., S. W. Cornelius, J. Smith et P. B. Baltes. "The Prototypical Meaning of Wisdom and Wise Persons." non publié (1990).
- Steele, C. M. "A Threat in the Air:: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist* 52, no. 6 (1997): 613-629.
- Sternberg, R. J. "A Balance Theory of Wisdom." *Review Of General Psychology* 2 (1998): 347-365.
- . "Costs of Expertise." in *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games*, édité par K. A. Ericsson, 347-354. Hillsdale, NJ: Lawrence Erlbaum Associates, 1996.
- . "Implicit Theories of Intelligence, Creativity, and Wisdom." *Journal of personality and social psychology* 49, no. 3 (1985): 607-627.
- Stichter, M. "Ethical Expertise: The Skill Model of Virtue." *Ethical Theory and Moral Practice* 10, no. 2 (2007): 183-194.
- Stukas, A. A et M. Snyder. "The Influence of Possible Future Interactions on First Encounters." non publié: La Trobe University, 2003.
- . "Targets' Awareness of Expectations and Behavioral Confirmation in Ongoing Interactions." *Journal of Experimental Social Psychology* 38 (2002): 31-40.
- Svartberg, K. et B. Forkman. "Personality Traits in the Domestic Dog (Canis Familiaris)." *Applied animal behaviour science* 79 (2002): 133-157.
- Tajfel, H. et J. C. Turner. "The Social Identity Theory of Intergroup Behavior." in *The Psychology of Intergroup Relations*, édité par W. G. Austin, 7-24. Chicago: Nelson-Hall, 1986.
- Taylor, S. E. et S. T. Fiske. "Point of View and Perceptions of Causality." *Journal of Personality and Social Psychology* 32, no. 3 (1975): 439-445.
- Ten Berge, M. A. et B. De Raad. "The Structure of Situations from a Personality Perspective." *European Journal of Personality* 16, no. 2 (2002): 81-102.
- Terman, Lewis M. "Some Paradoxes of Personality, or Muckraking in the Psychology of Character." *New England Monthly* 44 (1911): 371-374.

- Thorndike, Edward L. et R. S. Woodworth. "The Influence of Improvement in One Mental Function Upon the Efficiency of Other Functions." *Psychological Review* 8 (1991): 247-261.
- Tooby, J. et L. Cosmides. "The Psychological Foundations of Culture." *The adapted mind: Evolutionary psychology and the generation of culture* (1992): 19–136.
- Valliant, G. "Adaptation to Life." *Boston: Little, Brown* (1977).
- van Lieshout, C. F. M., G. J. T. Haselager et M. A. G. van Aken. "Generalizability of Personality Types and Subtypes across Age, Gender, and Methods of Assessment." *Biennial Meeting of the International Society for the Study of Behavioral Development, Berne, Switzerland* (1998).
- von Wright, G. H. *The Varieties of Goodness*: Kegan Paul, 1972.
- Vranas, Peter B. M. "The Indeterminacy Paradox: Character Evaluations and Human Psychology." non publié: The University of Michigan, 2004.
- . "Lack of Character: Personality and Moral Behavior." *Philosophical Review* 113, no. 2 (2004): 284-288.
- . "Moral Behavior: Cross-Situational Consistency and Predictability." non publié: The University of Michigan, 2006.
- Webber, Jonathan. "Virtue, Character and Situation." *Journal of Moral Philosophy* 3, no. 2 (2006): 195-216.
- Webber, Jonathan "Character, Common-Sense, and Expertise." *Ethical Theory and Moral Practice* publié online (2006).
- Welchman, Jennifer. "Virtue Ethics and Human Development: A Pragmatic Approach." in *Virtue Ethics, Old and New*, édité par S. M. Gardiner. Ithaca: Cornell University Press, 2005.
- Weyant, J. et R. D. Clark. "Dimes and Helping - Other Side of Coin." *Personality and Social Psychology Bulletin* 3, no. 1 (1977): 107-110.
- Williams, B. *Ethics and the Limits of Philosophy*. Cambridge, Mass: Harvard University Press 1985.
- Williamson, T. *Knowledge and Its Limits*: Oxford University Press, 2000.
- . "Philosophical 'Intuitions' and Scepticism About Judgement." *Dialectica* 58, no. 1 (2004): 109-153.
- Wright, R. *The Moral Animal: Evolutionary Psychology and Everyday Life*: Pantheon Books, 1994.
- Zagzebski, L. T. *Divine Motivation Theory*: Cambridge University Press, 2004.
- . *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*: Cambridge University Press, 1996.

Zanna, M. P., J. M. Olson et R. H. Fazio. "Attitude-Behavior Consistency - an Individual Difference Perspective." *Journal of Personality and Social Psychology* 38, no. 3 (1980): 432-440.

Zimbardo, P. G. "Transforming Experimental Research into Advocacy for Social Change." in *Applying Social Psychology: Implications for Research, Practice, and Training*, édité par M. Deutsch H. A. Hornstein. New York: Wiley, 1975.

Zuckerman, Marvin. *Psychobiology of Personality*. 2ème édition. New York, NY: Cambridge University Press, 2005.