



HAL
open science

Variables latentes et processus mentaux : une réflexion épistémologique et méthodologique

Hervé Guyon

► **To cite this version:**

Hervé Guyon. Variables latentes et processus mentaux : une réflexion épistémologique et méthodologique. Psychologie. Université Sorbonne Paris Cité, 2016. Français. NNT : 2016US-PCB015 . tel-01589014

HAL Id: tel-01589014

<https://theses.hal.science/tel-01589014>

Submitted on 18 Sep 2017

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 Descartes

Ecole Doctorale Cerveau Cognition Comportement (ED 158 3C)

CESP/INSERM U1018

Variables latentes et processus mentaux :
une réflexion épistémologique et
méthodologique.

Par Hervé GUYON

Thèse de doctorat de *Sciences Cognitives*

Dirigée par Bruno Falissard et Jacques Juhel

Présentée et soutenue publiquement le 29 avril 2016

Devant un jury composé de :

Véronique SEBILLE, Rapporteuse

Stéphane VAUTIER, Rapporteur

Jean-Luc KOP

Alain LEPLEGE

Résumé : Ma thèse développe deux parties. La première considère que la psychologie expérimentale doit clarifier son positionnement épistémologique pour clarifier la validation formelle de sa démarche, sans forcément devoir se référencer au cadre de la Science Physique. A partir d'une réflexion critique, je propose de décaler le cadre épistémologique en psychologie et de poser clairement un cadre pragmatique-réaliste. La thèse essentielle défendue dans ce travail est : 1/ les propriétés mentales doivent être comprises comme des phénomènes émergents, ce qui implique que leurs analyses ne peuvent se faire ni au niveau neuronal, ni au niveau de la dynamique interne de processus cognitifs, mais nécessairement au niveau de ces phénomènes émergents ; 2/ pour analyser les propriétés mentales comme formes émergentes, la psychométrie a besoin d'user de concepts qui sont en tension permanente entre une objectivité et une intersubjectivité ; en conséquence, la psychométrie doit affirmer une démarche pragmatiste-réaliste, en rupture avec l'empirisme-réaliste classique ; 3/ une approche pragmatiste-réaliste, basée entre autre sur l'abduction, permet de dépasser les contradictions pointées dans la littérature académique sur les propriétés mentales et leurs mesures ; 4/ un cadre de mesure de propriétés mentales par des variables latentes devient dès lors possible si ce cadre est compris lui aussi comme pragmatiste-réaliste ; 5/ mais ce recours au pragmatisme-réaliste renvoie en conséquence une critique à la fois des modèles avec variables latentes développés dans la littérature académique et les usages sociaux de ces modèles.

La seconde partie de ma thèse porte sur un cadre particulier de formalisation des variables latentes : le cadre formatif. Je développe des simulations Monte Carlo pour vérifier le spectre des paramètres permettant une mesure formative efficace dans le cadre d'un positionnement réaliste-empirique.

Title: Mental processes and latent variables: an epistemological and methodological reflection

Abstract: My thesis considers that experimental psychology must clarify its epistemological position to clarify the formal validation of its approach, without necessarily having to refer to the framework of Science Physics. From a critical reflection, I propose to shift the epistemological framework in psychology and clearly pose a pragmatic-realistic framework. The main thesis of this work is: 1 / mental properties must be understood as emerging phenomena, which implies that their analysis can not be done nor at the neuronal level, nor at the internal dynamics of cognitive processes, but necessarily at these emerging phenomena; 2 / to analyze the mental properties as emerging forms, psychometrics need to use concepts that are in permanent tension between objectivity and intersubjectivity; accordingly,

psychometrics must assert a pragmatic-realist approach, breaking with classical empiricism-realistic; 3 / a pragmatist-realistic approach, based among other things on the abduction, can overcome the contradictions pointed in the academic literature on mental properties and their measurements; 4 / a framework for measuring mental properties by latent variables becomes possible if the framework is also understood as a pragmatic-realist; 5 / but use realistic-pragmatic returns accordingly critical of both models with latent variables developed in the academic literature and the social uses of these models.

The second part of my thesis focuses on a specific part of formalization of latent variables: the formative model. I develop Monte Carlo simulations to check the range of parameters for efficient formative measure as part of a realistic-empirical positioning.

Mots clés (français) : Emergence, Epistémologie, Modèles de mesures formatives, Pragmatisme, Processus mentaux, Simulations Monte Carlo, Variables latentes

Keywords: Emergence, Epistemology, Formative measurement models, Latent variable, Mental processes, Monte Carlo simulations, Pragmatism

A Camille

Remerciements

J'espère que ces remerciements seront compris non pas comme une obligation rituelle, mais comme des sentiments sincères après ce parcours intellectuel qui a engagé bien plus que ma personne.

Bien évidemment, et ceci n'est pas une volonté de suivre un protocole normatif, je remercie en premier lieu mes deux directeurs de thèse, Bruno Falissard et Jacques Juhel. Sans l'interaction que nous avons eue je ne serais jamais arrivé à finaliser ce texte dans sa forme actuelle. Ils ont, chacun d'eux avec leurs qualités propres, stimulé ma recherche d'une façon positive et encourageante. Je ne sais pas si le résultat est à la hauteur de leurs ambitions. Mais ce fut pour moi, et je l'espère pour eux, une belle aventure intellectuelle et humaine qui va se prolonger j'en suis sûr.

Je remercie aussi les différent.e.s collègues qui ont partagé avec moi ma réflexion durant ces années de réflexion (Dominique Martin, Georges Oppenheim, Alexandra Rouquette, Dominique Roux, Mouloud Tensaout, ...), jusqu'à devenir rapporteur.e.s ou membres de mon jury de thèse pour certain.e.s (Jean-Luc Kop, Véronique Sébille, Stéphane Vautier).

Je remercie chaleureusement Alain Leplège d'avoir accepté d'être membre de mon jury de thèse sans avoir eu l'opportunité de partager sa génèse.

Ces échanges intellectuels ont aussi rempli ma vie sociale. Que mes ami.e.s soient chaleureusement remercié.e.s de leur apport, non négligeable, et de leurs amitiés inébranlables : Hélène, Chloé, Laurence, Caroline, Fabrice, Grégoire, Cécile, Olivier, Noémie, ... (que les oublié.e.s m'excusent). Avec eux aussi l'aventure humaine et intellectuelle va continuer... avec des discussions à suivre dans nos soirées agrémentées de bonnes bouteilles.

Ma dysorthographe et dyslexie a contraint mon entourage familial et amical à devoir participer à ce travail par la tâche ingrate de relecture d'un texte relativement hermétique. S'il reste des fautes, veuillez excuser Adèle, Anne-Marie, Chloé, Gérard, Hélène, Laurence, Morgane, Muriel. Je n'en suis que le seul responsable. Cependant, j'ai toujours adapté mon positionnement au sujet de l'orthographe relativement à mes compétences : considérer que les fautes n'étaient en soi pas un problème si elles n'altéraient pas le sens d'un texte. J'espère donc sincèrement que les fautes qui resteraient dans ce texte n'altéreront pas le sens de mon propos.

Je remercie aussi chaleureusement Sanga, mon chat, de sa participation active. Lorsqu'il venait se poser sur mes genoux, il avait un air intéressé très encourageant en lisant ce que j'écrivais sur l'écran de mon ordinateur. Sa participation active par son usage intempestif du clavier ou de la souris de mon ordinateur a parfois généré la rédaction de nouveaux concepts, que je n'ai pas repris car la substance théorique m'est restée obscure. Mais il a surtout veillé à ma santé en renversant sans vergogne ma tasse de café quand il semblait considérer que j'en avais déjà assez bu.

Mais aucune de ces personnes n'a eu à me subir comme Muriel. Je fus par alternance obsessionnel (préoccupé uniquement par ma thèse), autiste (insensible à mon environnement affectif), bipolaire (alternant phase dépressives et surexcitation). Muriel a accepté le rôle de la compagne attentive, réduisant mon féminisme, déjà plus théorique que pratique dans la vie normale, qu'à de vagues déclarations de principe pendant de longues semaines. Pire, Muriel a dû endosser le rôle de la contradiction intellectuelle, devant abandonner, lorsque je daignais sortir de ma bulle, ses propres travaux pour devoir discuter du sujet de ma thèse. Merci à elle d'être qui elle est.

Table des matières

<u>REMERCIEMENTS</u>	7
<u>TABLE DES ILLUSTRATIONS</u>	14
<u>AVERTISSEMENT</u>	16
<u>PREAMBULE</u>	19
<u>INTRODUCTION</u>	21
<u>LA SEMANTIQUE</u>	23
<u>L'OBJET DE LA THESE</u>	25
<u>LA THESE DEVELOPEE</u>	28
<u>LES DEUX PARTIES DE LA THESE</u>	30
<u>LA FORME DE LA THESE</u>	31
<u>UNE DISCUSSION</u>	32
<u>STRUCTURE DE LA THESE</u>	34
<u>LA REFLEXIVITE DU REDACTEUR</u>	35
<u>PREMIERE PARTIE : REFLEXION EPISTEMOLOGIQUE SUR LES PROPRIETES MENTALES ET VARIABLES LATENTES</u>	39
<u>1. ONTOLOGIE, EPISTEMOLOGIE ET PROPRIETES MENTALES</u>	41
<u>1.1. LA NATURE ONTOLOGIQUE DES PROPRIETES MENTALES</u>	41
<u>1.1.1. LE CADRE PHILOSOPHIQUE DE LA DISCUSSION ACADEMIQUE</u>	42
<u>1.1.2. LA NATURE PHYSIQUE DES PROPRIETES MENTALES</u>	47
<u>1.1.3. LA THEORIE DES SYSTEMES COMPLEXES ASSOCIEE AU CERVEAU</u>	54
<u>1.1.4. LA PROPRIETE MENTALE COMME PHENOMENE EMERGENT ?</u>	62
<u>1.1.5. LA PROPRIETE MENTALE COMME QUALITE EMERGENTE ?</u>	72
<u>1.1.6. CONCLUSION SUR L'ONTOLOGIE DES PROPRIETES MENTALES</u>	81
<u>1.2. LE CADRE EPISTEMOLOGIQUE</u>	86

1.2.1	LE PROBLEME DE LA CONCEPTUALISATION DES PROPRIETES MENTALES	86
1.2.2	RUPTURE EPISTEMOLOGIQUE AVEC L'EMPIRISME-REALISTE	91
1.2.3	LE CONCEPT DE CONSTRUIT	104
1.2.4	CONCLUSION SUR LE CADRE EPISTEMOLOGIQUE	109
2	LA MESURE DES PROPRIETES MENTALES	115
2.1	LE PROBLEME DE LA MESURE EN PSYCHOLOGIE	115
2.1.1	SCIENCE PATHOLOGIQUE ?	115
2.1.2	LE CADRE FORMEL CLASSIQUE DE LA MESURE	117
2.1.3	L'APPROCHE DE STEVENS	118
2.1.4	LA THEORIE REPRESENTATIONNISTE DE LA MESURE	118
2.1.5	LES MESURES CONJOINTES	120
2.1.6	LATENT VARIABLE THEORY	124
2.1.7	LE MODELE DE RASCH	129
2.1.8	CONCLUSION	136
2.2	MESURER QUOI ?	140
2.2.1	L'OBJECTIVITE DE LA PROPRIETE MENTALE ?	140
2.2.2	CATEGORISATION DE LA PROPRIETE MENTALE ?	145
2.2.3	UNIDIMENSIONALITE DE LA PROPRIETE MENTALE ?	148
2.2.4	INDEPENDANCE LOCALE ?	150
2.2.5	CONCLUSION	153
2.3	COMMENT MESURER PRAGMATIQUEMENT ?	154
2.3.1	QUE VEUT DIRE MESURER UNE PROPRIETE MENTALE ?	154
2.3.2	RETOUR A LA RTM	157
2.3.3	QUE VEUT DIRE ATTRIBUER UNE VALEUR NUMERIQUE ?	159
2.3.4	COMMENT ATTRIBUER UNE VALEUR NUMERIQUE ?	162
2.3.5	VALIDITE DES MESURES SUBJECTIVES ?	166
2.3.6	L'EXEMPLE DE LA DOULEUR	169
2.4	LA PRATIQUE DE LA MESURE	177
2.4.1	LE PROBLEME DU MODELE STATISTIQUE EN PSYCHOLOGIE	177
2.4.2	LE PROBLEME DES PROBABILITES	186
2.4.3	LE PROBLEME DE LA FALSIFIABILITE	191
2.4.4	LE PROBLEME DE « L'INDIVIDU MOYEN » VS « LE SUJET »	197
2.4.5	MODELES AVEC VARIABILITE INTRA	199

2.4.6	ECHELLE DE MESURE ET TESTS	203
2.4.7	CONCLUSION	212
2.5	CRITIQUE DE L'USAGE SOCIAL DE LA MESURE	215
2.5.1	IMPERATIF QUANTITATIF ?	215
2.5.2	MODELE ET SINGULARITE	220
2.5.3	ROLE DES MESURES	223
2.5.4	EVALUER ET COMPRENDRE	228
2.5.5	L'EXEMPLE DE L'INTELLIGENCE	232
2.5.6	CRITIQUE DE LA SCIENCE	239
2.5.7	CONCLUSION	243
3	CONCLUSION DE LA PARTIE 1	245
3.1	LA THESE DEFENDUE	245
3.1.1	MA THESE	245
3.1.2	LE PRAGMATISME COMME METHODE	246
3.2	UN PRAGMATISME USUEL ?	248
3.2.1	LE PRAGMATISME DANS LA LITTERATURE	248
3.2.2	LE GOLDILOCKS DE MARKUS ET BORSBOOM	250
3.2.3	CONCLUSION	252
3.3	SOFT MEASUREMENT ?	253
3.3.1	LES BASES DU <i>SOFT MEASUREMENT</i>	253
3.3.2	LA DEMARCHE ACTUELLE DU <i>SOFT MEASUREMENT</i>	255
3.3.3	CONVERGENCE ET DIVERGENCE AVEC LE <i>SOFT MEASUREMENT</i>	258
3.4	CONCLUSION	259
	PARIE 2 : REFLEXION METHODOLOGIQUE SUR LA MESURE FORMATIVE D'UNE PROPRIETE MENTALE	261
	PREMBULE A LA SECONDE PARTIE	263
4	INTRODUCTION AUX MESURES FORMATIVES	267
4.1	LA NATURE DISCUTABLE D'UNE MESURE FORMATIVE	268
4.1.1	LES MESURES FORMATIVES RECENTMENT INTRODUITES	268

4.1.2	COMMENT SE FORMALISE UN CONSTRUIT FORMATIF	269
4.1.3	LA NATURE ONTOLOGIQUE DES CONSTRUIES FORMATIFS DU POINT DE VUE EMPIRISTE-REALISTE	270
4.1.4	LES MESURES FORMATIVES D'UN POINT DE VUE PRAGMATIQUE-REALISTE	272
4.1.5	LES PROBLEMES METHODOLOGIQUES	277
4.1.6	CONCLUSION	279
4.2	CONTRAINTES ET MENACES SUR LES MESURES FORMATIVES	281
4.2.1	HYPOTHESES FORMELLES ENTRE LE CONSTRUIT ET SES INDICATEURS FORMATIFS	281
4.2.2	IDENTIFICATION DU MODELE DE MESURE FORMATIF	284
4.2.3	LE PROCESSUS DE VALIDATION D'UN MODELE DE MESURE FORMATIF	287
4.2.4	CONCLUSION	301
4.3	SIMULATIONS	303
4.3.1	LE CADRE DES SIMULATIONS	303
4.3.2	LE MODELE THEORIQUE	303
4.3.3	LES LACUNES DES SIMULATIONS DANS LA LITTERATURE	304
4.3.4	DESIGN DES SIMULATIONS MONTE CARLO	308
4.3.5	CRITERES	313
4.3.6	LE CODE SIMULE	317
4.3.7	LES SIMULATIONS	319
4.3.8	LES RESULTATS	320
4.4	L'INTERPRETATION CONFONDANTE	321
4.4.1	L'INTERPRETATION CONFONDANTE	321
4.4.2	SIMULATIONS	323
4.4.3	RESULTATS SUR L'INSTABILITE DES PARAMETRES FORMATIFS	324
4.4.4	RESULTATS SUR LE SCORE	327
4.4.5	VALIDITE DU MODELE STRUCTUREL	329
4.4.6	CONCLUSION SUR LE MODELE DE MESURE FORMATIF	330
4.5	UNE LATENTE FORMATIVE EST-ELLE REFLECTIVE ?	332
4.5.1	LES ARGUMENTS THEORIQUES	332
4.5.2	SIMULATIONS	336
4.5.3	RESULTATS	337
4.5.4	CONCLUSION	340
4.6	DETECTION DES MAUVAISES SPECIFICATIONS	341
4.6.1	DETECTION DU NON RESPECT DES CONTRAINTES DE PROPORTIONALITE	341
4.6.2	DETECION DE LA NON EXHAUSTIVITE DES INDICATEURS FORMATIFS	342

4.6.3	CONCLUSION	343
4.7	PLS-SEM	344
4.7.1	PLS-SEM	344
4.7.2	SIMULATIONS	345
4.7.3	COMPARAISONS PLS-SEM ET CB-SEM POUR LE CRITERE DES INDICATEURS FORMATIFS	345
4.7.4	COMPARAISON CB-SEM ET PLS-SEM SUR LE SCORE ESTIME	346
4.7.5	COMPARAISON DE CB-SEM ET PLS-SEM SUR LE MODELE STRUCTUREL	348
4.7.6	CONCLUSION	349
4.8	DISCUSSION GENERALE	351
4.8.1	LES RESULTATS DE NOS SIMULATIONS	351
4.8.2	LE CADRE D'UTILISATION D'UN MODELE DE MESURE FORMATIF	351
4.8.3	L'EXEMPLE DE KIM ET AL. 2011	353
5	CONCLUSIONS DE LA PARTIE 2	355
5.1	MODELES FORMATIFS A EVITER	355
5.2	CB-SEM OU PLS-SEM	356
5.3	RECOMMANDATIONS	357
	CONCLUSION GENERALE	361
	LE FETICHISME DE LA MESURE EN PSYCHOLOGIE	363
	POUR UNE MESURE OBJECTIVE DES MESURES SUBJECTIVES	364
	LE PRAGMATISME COMME METHODE	366
	LE PRAGMATISME COMME CRITIQUE DE LA PRATIQUE DE LA SCIENCE	367
	PRAGMATISME ET CRITIQUE SOCIALE	369
	BIBLIOGRAPHIE	373
	ANNEXES 1 : CODE R DES SIMULATIONS	417

Table des illustrations

Figure 1 : Modèle de mesure d'une propriété mentale	25
Figure 2 : Modèle formatif	30
Figure 3 : Cadre d'une relation « réaliste » entre un phénomène latent et ses manifestations visibles	44
Figure 4 : Cadre d'une relation « constructiviste/instrumentaliste » entre des indicateurs et un phénomène	44
Figure 5 : Représentation d'attracteurs	56
Figure 6 : Schématisation d'un modèle hiérarchique avec variables latentes	57
Figure 7 : Approche par variables latentes de la dépression	59
Figure 8 : Approche <i>Network Analysis</i> de la dépression	60
Figure 9 : Figure de la propriété émergente d'un <i>Neural Networks</i>	75
Figure 10 : Figure de la propriété émergente d'un <i>Neural Networks</i> inscrit dans une praxis ..	77
Figure 11 : Séquence du processus de recherche	101
Figure 12 : L'exemple de la douleur	170
Figure 13 : Schéma différenciant Propriété-Concept-Variable	182
Figure 14 : Schéma du processus d'une approche quantitative	183
Figure 15 : Processus de la construction et de la validation d'un test (<i>goldilocks</i>)	251
Figure 16 : Un modèle de mesure formatif d'un construit de premier ordre	268
Figure 17 : Un modèle de mesure réflectif d'un construit de premier ordre	268
Figure 18 : Modèle MIMIC	273
Figure 19 : Variable latente liée à deux autres variables latentes	273
Figure 20 : a) Modèle estimé avec le lien entre ETA et Y1 contraint ; b) Modèle estimé avec le lien entre X1 et ETA contraint	285
Figure 21 : Modèle de base a) estimé avec CB-SEM * ; b) estimé avec PLS-SEM **	289
Figure 22 : a) Modèle théorique ; b) Modèle mal spécifié estimé avec CB-SEM * ; c) Modèle mal spécifié estimé avec PLS-SEM **	291
Figure 23 : Modèle sans X1 a) estimé avec CB-SEM * ; b) estimé avec PLS-SEM **	292
Figure 24 : Modèle avec une erreur de mesure a) estimé avec CB-SEM * ; b) estimé avec PLS-SEM **	293
Figure 25 : a) Modèle théorique avec colinéarité des X_i ; b) estimé avec CB-SEM ** ; c) estimé avec PLS-SEM **	295

Figure 26 : a) Modèle théorique avec une liaison fallacieuse ; b) estimé avec CB-SEM ** ; c) estimé avec PLS-SEM **	297
Figure 27 : Modèle de base avec Y3 et Y4 a) estimé avec CB-SEM * ; b) estimé avec PLS-SEM **	299
Figure 28 : Construit endogène avec des indicateurs formatifs qui pose le paradoxe de la mauvaise spécification du modèle de mesure.	300
Figure 31 : Construit avec des indicateurs formatifs endogènes qui pose le paradoxe de la variable latente formative non directement influencée par d'autres variables latentes.	300
Figure 30 : Modélisation d'une variable latente formative	301
Figure 31 : Modèle MIMIC	308
Figure 32: MODELE 1, Variable latente formative η reliée à deux latentes réfléchives η_1 et η_2	309
Figure 33 : MODELE 2, Variable latente formative η reliée à quatre latentes réfléchives η_1 à η_4	310
Figure 34 : Modèle MIMIC avec le non-respect des règles de proportionnalité.	312
Figure 35 : Modèle de mesure formatifs vs modèle de mesure réfléchif d'une même latente 333	
Figure 36 : Biais de SMC en fonction des valeurs des paramètres λ_1* λ_2	335

Avertissement

L'Université Paris Descartes n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

« Ce qui m'a frappé, c'est la manie qu'ont les médecins de la santé publique, et parmi eux particulièrement les psychiatres, de toujours tout mettre sous forme de chiffres et de toujours avoir des échelles pour tout mesurer. Ils ont des échelles de dépression, des échelles d'indépendance, des échelles de tout ce qu'on veut ».

A. Gramin (*in* Desrosières, 2008b, p. 183)

Préambule

L'origine de cette thèse s'inscrit dans une trajectoire suffisamment atypique pour l'expliquer. A un âge (49-53 ans, 49 ans était l'âge lors de mon inscription en thèse, 53 ans l'âge lors de ma soutenance) où la retraite est plus proche que le début de ma vie professionnelle (même si la date de ma retraite s'éloigne au gré des réformes successives), avec un statut social reconnu (enseignant chercheur titulaire dans une université parisienne), je me dois d'expliquer la raison de cette thèse.

J'ai commencé à travailler sur les variables latentes en 1993, lorsque je fus recruté comme ingénieur statisticien dans un institut d'études de marché, essentiellement sur des modèles d'équations structurelles (SEM). De façon surprenante pour ceux qui connaissent les SEM utilisés à cette époque, j'ai travaillé dès 1993 sur l'approche dite *covariance based* (LISREL que je nommerai CB-SEM comme la littérature actuelle tend à le faire) et l'approche *partial least square* (PLS-SEM ou PLSPM, avec le programme de Lohmöller sur DOS, lvpls1.8.exe, dont je ne sais plus comment j'avais obtenu une copie). Suite à une thèse en science de gestion (2005) je suis devenu enseignant chercheur. Mon cœur de compétence étant méthodologique, j'ai essentiellement travaillé durant toutes ces années antérieures et postérieures à ma thèse sur les SEM et autres modèles quantitatifs (analyse conjointe entre autres), comme en témoignent mes quelques publications académiques. Ces travaux de recherche usant des SEM me renvoyaient à des questions qui exigeaient que je prenne le temps de les creuser, et donc de m'échapper de la pression que subit un enseignant chercheur du devoir de résultats immédiats (la pression bibliométrique). Une thèse m'offrait la possibilité de rédiger dans le sens que je le souhaitais, sans devoir répondre aux impératifs de la ligne éditoriale d'une revue ou de reviewers. Bref, une thèse me donnait les moyens temporels et matériels d'aller au bout de mes questionnements.

Introduction

Le titre de cette thèse se voulait modeste à son amorce (septembre 2012) : « *Variables latentes et processus mentaux : une réflexion épistémologique et méthodologique*. ». Si j'ai utilisé le terme « réflexion » c'est que je n'avais aucune idée du cheminement intellectuel que j'allais effectuer pour répondre interrogations sur l'usage des modèles avec variables latentes auxquelles je me sentais confronté depuis des années. J'ai découvert au cours de ma thèse, dans des revues académiques et livres, des discussions que j'ignorais et qui renvoyaient à mes propres interrogations. In fine, mon travail est d'une modestie relative car cette thèse couvre des champs très différents et très vastes, et se veut une réponse personnelle à cet ensemble de discussions autour de la légitimité d'utiliser des modèles avec variables latentes en psychologie.

La sémantique

Avant de poursuivre je clarifie la sémantique utilisée par la suite.

1/ j'userai du terme générique de « **propriété mentale** » pour caractériser ce que d'autres auteurs nomment « activités mentales », « états psychologiques », « trait latents », « états émotionnels », « dimension psychologique », (je cite les termes trouvés, par les articles utilisant des modèles avec variables latentes sur les deux dernières années dans *Psychologie Française*). Le titre de ma thèse renvoyait initialement au terme de « processus mental ». Or, comme je le discuterai, nous ne cherchons pas à mesurer¹ un *processus* avec une variable latente mais bien une **propriété**. Le terme « **attribut** » (parfois utilisé dans la littérature) aurait pu être utilisé, mais il a un sens plus « physique » et « statique » que « propriété », or mon développement cherchera à caractériser une propriété mentale comme une réalité matérielle mais sans référent physique précis. J'utiliserai parfois le terme « attribut » lorsque je me référerai à d'autres auteurs pour reprendre la terminologie qu'eux ont employée. De même, j'use du terme « **mental** » plutôt que « psychologique » car ce dernier me paraît glisser vers une caractérisation plus cadrée que le terme de « mental » qui lui renvoie implicitement au cerveau sans plus de précision.

2/ j'userai parfois du terme « **état mental** » pour caractériser une forme particulière d'une propriété mentale. Ainsi je parlerai de l'anxiété par exemple comme une « propriété

¹ Le terme « mesure » est à prendre ici au sens large et non au sens d'une métrique attribuée. Je reviendrai longuement sur « la mesure » par la suite

mentale », mais pour une personne donnée, cette propriété se caractérise par un « état mental » précis.

3/ j’userai parfois du terme de « **construit** » car c’est le terme générique utilisé dans la littérature académique (« *construct* ») pour caractériser une propriété mentale avec des variables latentes, mais ce terme « construit » peut porter à confusion comme je l’expliquerai dans une section dédiée à ce concept ; néanmoins son caractère confus est en réalité parfois utile.

Si par mégarde il restait dans ce texte d’autres termes que ceux-ci, je m’excuse d’avance. Mais l’unité sémantique proposée ci-dessus découle des conclusions de ma réflexion, et donc elle est postérieure à bien des parties rédigées antérieurement à ce choix. Il y a donc peut être quelques scories de rédactions antérieures ayant échappé à ma vigilance.

Sur le plan de l’épistémologie, je me référerai souvent à « l’empirisme-réaliste », au « constructivisme » et au « pragmatisme », termes explicités au cours de mon développement. Pour faciliter la lecture de cette thèse, je donne ici des définitions succinctes qui ne prétendent pas répondre à un glossaire de philosophie :

L’empirisme-réaliste : cadre épistémologique qui considère que les objets étudiés sont réels et extérieurs au chercheur ; l’objectif de la démarche scientifique est de découvrir/caractériser ces objets.

Le constructivisme : cadre épistémologique qui considère que les objets étudiés sont avant tout construits par le chercheur sans pouvoir affirmer une réalité externe au chercheur.

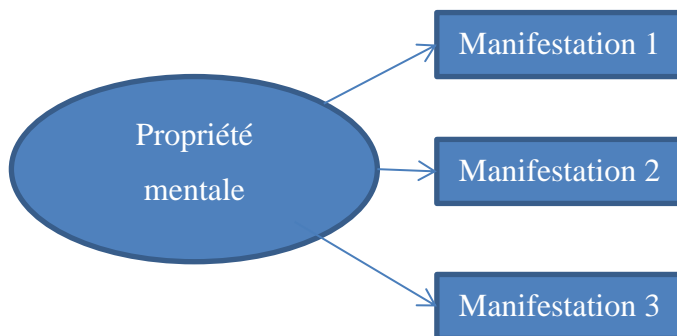
Le pragmatisme : cadre épistémologique qui considère que les objets étudiés sont réels (à la différence du constructivisme), mais leur connaissance est liée à une pratique (il n’y a pas de « vérité », ou de « quête de cette vérité », dans cette connaissance à la différence de l’empirisme-réaliste).

Je me référerai souvent au **pragmatisme-réaliste**, qui met l’accent sur la nature « réelle » des objets étudiés, se différenciant d’un pragmatisme glissant vers le relativisme.

L'objet de la thèse

Un modèle utilisant des variables latentes pour inférer des propriétés mentales considère qu'il y a des phénomènes observés (les items d'un questionnaire ou autres phénomènes observables) et des phénomènes inobservables (les propriétés mentales comme l'anxiété, l'intelligence, la dépression, ...). L'objectif est de pouvoir « mesurer² » la propriété mentale à partir des manifestations visibles qu'elle engendre ; c'est-à-dire construire un modèle avec des variables, certaines étant directement mesurées (les phénomènes observés, comme des réponses à des items par exemple), les autres (les variables latentes) mesurant des propriétés mentales latentes (inobservables directement).

FIGURE 1 : MODELE DE MESURE D'UNE PROPRIETE MENTALE



Ce modèle peut se représenter par un formalisme mathématique, où les X_i représentent sous forme de variables les manifestations observables (par exemple des réponses à des items), appelées « variables manifestes » ; les λ_i représentent les liens entre les variables manifestes et la variable latente (le plus souvent linéaires, à analyser comme des coefficients de régression classique); les ε_i sont des « erreurs de mesure » (on considère que le modèle n'est pas « exact »), et η la variable latente représentant la propriété mentale :

$$X_i = \lambda_i \times \eta + \varepsilon_i \quad (1)$$

Que l'on peut écrire :

$$E(X|\eta) = \lambda\eta + \epsilon \quad (2)$$

Pour résumer et sans rentrer dans les détails, l'objectif est d'estimer les λ_i ou d'estimer η .

² Le terme « mesure » est à prendre ici au sens large et non au sens d'une métrique attribuée. Je reviendrai longuement sur « la mesure » par la suite

Le premier modèle de variable latente remonte à Spearman (1904). Depuis, beaucoup de modèles avec variables latentes existent et ces modèles connaissent une popularité dans toutes les *Sciences Humaines et Sociales* (SHS) rarement égalée. Borsboom (2008b) résume ainsi, bien que considérant que bien d'autres combinaisons soient possibles, l'ensemble des modèles potentiels avec variables latentes : « *If the observed and latent variables are both continuous, then the resulting model is called a factor model (Jöreskog, 1971; Lawley & Maxwell, 1963; Bollen, 1989); if the observed variable is categorical and the latent variable is continuous, then we have an Item Response Theory (IRT) model (Rasch, 1960; Birnbaum, 1968; Hambleton & Swaminathan, 1985; Embretson & Reise, 2000; Sijtsma & Molenaar, 2002); if the observed and latent variables are both categorical, the resulting model is known as a latent class model (Lazarsfeld & Henry, 1968; Goodman, 1974); and if the observed variable is continuous while the latent variable is categorical, then we get a mixture model (McLachlan & Peel, 2000), which upon appropriate distributional assumptions becomes a latent profile model (Lazarsfeld & Henry, 1968; Bartholomew, 1987).* » (p.25).

Déjà en 1986, Bentler (1986) notait que ces modèles avaient eu une popularité rare en recherche en psychologie, conséquence de la possibilité de traiter des problématiques peu gérables autrement (« *Unlike many other developments in theoretical psychometrics, this methodology spread from the methodology laboratory into the research laboratory with unusual rapidity. One major reason for such a quick and pervasive penetration of many fields of psychological and social science research is that the methods permit researchers to effectively study substantive problems that could not easily be investigated using alternative approaches.* », p. 35). Le développement de ces modèles s'est fait dans deux sens : un développement très important des méthodes statistiques ; un développement très important des applications dans tous les champs des SHS. A ce sujet, Osborne (2010) discute des challenges de la psychologie quantitative au XXIème siècle, et pour lui au-delà des évolutions statistiques et algorithmiques, il faut toujours questionner les fondamentaux. Pour Osborne, le fait que les tests³ donnent des structures stables semble montrer que ces données sont issues de quelque chose, que les items d'un test sont conséquents d'un attribut réel. Cependant, au-delà de l'affirmation à partir de résultats empiriques (« *puisque cela marche, c'est que cela doit être juste* » ?), quels fondements scientifiques peut-on donner aux modèles avec variables latentes en dehors d'une fonction purement instrumentaliste ?

³ Au sens de « la théorie du test » en psychologie, c'est-à-dire les mesures effectuées par des « tests psychologiques » que j'élargirai par la suite aux « échelles de mesure »

Depuis de nombreuses années, une discussion s'est développée pour clarifier épistémologiquement ces modèles. Nous pouvons considérer que les articles de Borsboom, Mellenbergh, & van Heerden (2003, 2004), très souvent cités dans la littérature contemporaine, ont remis au premier plan une réflexion sur ces modèles. Depuis, la discussion est vive entre d'un côté des courants considérant ces modèles comme incontournables, et de l'autre des chercheurs les considérant comme basés sur des hypothèses erronées. *Journal of Personality Assessment* en 2005 (P. Barrett, 2005; Kagan, 2005; Maraun & Peters, 2005; McGrath, 2005a, 2005b, p. 2005) ; *Measurement: Interdisciplinary Research & Perspective* en 2008 (P. Barrett, 2008; Borsboom, 2008a, 2008b; Cervone & Caldwell, 2008; Haig, 2008a; Haig & Borsboom, 2008; Hood, 2008; Howell, 2008; Kane, 2008; Kyngdon, 2008c; Maraun & Halpin, 2008; Markus, 2008a, 2008b; Michell, 2008b, 2008c; Mislavy, 2008; Rupp, 2008) et 2012 (Borsboom, 2012; Hood, 2012; Lissitz & Califo, 2012; Maraun, 2012; Markus, 2012; Newton, 2012) ; *New Ideas in Psychology* en 2013 (Lovasz & Slaney, 2013; Maraun & Gabriel, 2013; Markus & Borsboom, 2013b; Michell, 2013; Schmittmann et al., 2013; Slaney & Racine, 2013a, 2013b) ; *Journal of Educational Measurement* en 2013 (Borsboom & Markus, 2013; Brennan, 2013; Haertel, 2013; Kane, 2013a; Moss, 2013; Newton, 2013; Sireci, 2013) ont par exemple fait des numéros spéciaux ; ainsi qu'un ouvrage collectif souvent cité (Lissitz, 2009), autour de cette discussion dans laquelle aucun consensus ne semble se dégager. De même une discussion sur la nature réelle des construits mentaux a été impulsée à partir de l'article de Kievit et al. (2011a) dans *Psychological Inquiry* (Bagozzi, 2011a; L. F. Barrett, 2011; Berkman & Lieberman, 2011; Burnston, Sheredos, & Bechtel, 2011; Haig, 2011; Kievit et al., 2011b; Vul, 2011).

Une forme de consensus se dégage cependant dans cette littérature, à de rares exceptions près, sur le cadre formel des modèles avec variables latentes. Dans un travail avec variables latentes, le chercheur se positionne le plus souvent, plus ou moins implicitement, dans le cadre de l'empirisme-réaliste ; c'est-à-dire considère que la propriété mentale est un objet réel extérieur au chercheur à analyser (« *des entités réelles, en attente d'être découvertes* » Juhel & Gilles, 2011). Ce cadre de référence empiriste-réaliste renvoie à la physique en tant que science de référence. On est obligé d'avoir dès lors recours au protocole et au formalisme de la Science Physique pour valider ses hypothèses. Sur la base de ce cadre épistémologique, la psychologie expérimentale utilisant des modèles avec variables latentes échoue concrètement à valider sa démarche et ses hypothèses, comme je le discuterai dans mon texte : la réalité d'une propriété mentale ne peut pas être caractérisée physiquement ; sa nature quantitative ne peut pas être démontrée ; la validation formelle d'un modèle ne peut pas être effectuée. Dès

lors, ne faut-il pas arrêter d'utiliser les variables latentes ? Bien des auteurs, quoique minoritaires, développent de telles thèses contre la psychométrie actuelle : Maraun parle de « *hérésie* », Michell parle de « *Science pathologique* », Vautier parle de « *fétichisme* » (je résume par les mots phares de ces auteurs car leurs propos ne sont pas toujours aussi radicaux, mais ils sous-tendent cette critique de façon récurrente).

La discussion menée dans cette thèse peut se formaliser ainsi : **Les modèles avec variables latentes ont-ils une légitimité scientifique ?**

La thèse développée

Il existe une critique argumentée sur la psychométrie. Bien que pertinente sur de nombreux points, cette critique ne me convainc pas. Notamment le cadre épistémologique de la critique est en général le même que celui des hérauts des modèles avec variables latentes, sans que ne soit posée son efficacité en psychologie : le cadre empiriste-réaliste.

A partir des articles posant le cadre de l'usage des variables latentes et des articles critiques d'une telle formalisation, j'ai été amené à questionner ce cadre épistémologique utilisé dans la quasi-totalité des travaux en psychologie, autant par les défenseurs que par les opposants à ces méthodes. Bien des philosophes des sciences questionnent « l'unité » de la science (voir par exemple Feyerabend, 1979/2005). Dans leur sillage, je considère que la psychologie expérimentale doit clarifier son positionnement épistémologique pour clarifier la validation formelle de sa démarche, sans forcément devoir se référer au cadre de la Science Physique. A partir d'une réflexion critique, j'en suis arrivé à la conviction qu'il fallait en quelque sorte décaler le cadre épistémologique en psychologie et poser clairement un cadre pragmatique-réaliste. A partir de ce cadre épistémologique, les modèles avec variables latentes m'apparaissent scientifiquement légitimes... mais avec en conséquence une critique de leur usage actuel. Ce sera l'objet de tout mon développement dans la première partie de cette thèse.

L'objectif de ma discussion épistémologique n'est pas de prendre une position partisane sur les sciences en général, mais de discuter d'un positionnement clarifié, et donc prendre position, par rapport à un objet particulier : la validité scientifique des propriétés mentales mesurées par des variables latentes. Adopter ce parti pris permet de contourner la question de l'unification de la démarche scientifique et de laisser ouverte la possibilité que la démarche

fondamentale par rapport à ces objets puisse se différencier de démarches dans d'autres champs disciplinaires.

In fine, ce travail se veut une réponse épistémologique et méthodologique sur la capacité des variables latentes à apporter du sens sur des propriétés mentales. La thèse essentielle défendue dans ce travail est : 1/ les propriétés mentales doivent être comprises comme des phénomènes émergents, ce qui implique que leurs analyses ne peuvent se faire ni au niveau neuronal, ni au niveau de la dynamique interne de processus cognitifs, mais nécessairement au niveau de ces phénomènes émergents ; 2/ pour analyser les propriétés mentales comme formes émergentes, la psychométrie a besoin d'user de concepts qui sont en tensions permanentes entre une objectivité et une intersubjectivité ; en conséquence, la psychométrie doit affirmer une démarche pragmatiste-réaliste, en rupture avec l'empirisme-réaliste classique ; 3/ une approche pragmatiste-réaliste, basée entre autre sur l'abduction, permet de dépasser les contradictions pointées dans la littérature académique sur les propriétés mentales et leurs mesures ; 4/ un cadre de mesure de propriétés mentales par des variables latentes devient dès lors possible si ce cadre est compris lui aussi comme pragmatiste-réaliste ; 5/ mais ce recours au pragmatisme-réaliste renvoie en conséquence une critique à la fois des modèles avec variables latentes développés dans la littérature académique et les usages sociaux de ces modèles.

Kuhn (1970) discute des révolutions paradigmatiques en Science. L'adhésion à de nouveaux paradigmes est souvent effectuée par des jeunes chercheurs, laissant la génération plus ancienne défendre dans un combat vain d'anciens paradigmes et s'enliser dans leur impasse. La popularité d'approches récentes comme les *Network Analysis*, dont je discuterai, peut être comprise comme une rupture avec le paradigme dominant et amorcerait une telle révolution paradigmatique en psychométrie. Ma thèse peut donc être comprise comme une résistance (vaine) des modèles avec variables latentes contre une évolution/révolution nécessaire de la psychométrie. J'espère que ce texte montrera le contraire. Si je défends l'usage des modèles avec variables latentes, et donc suis critique de la critique des modèles avec variables latentes, j'adhère pourtant à bien de ces critiques développées dans la littérature contre le paradigme dominant sur les variables latentes. L'objet de ma thèse n'est donc absolument pas de défendre ce paradigme dominant en psychométrie. Mais je n'adhère pas non plus à la critique dominante (Michell, Maraun ou Vautier par exemple, autant que l'on puisse les regrouper) de ce paradigme car à mes yeux leurs critiques s'inscrivent dans un cadre épistémologique tout aussi critiquable, même si cette critique me paraît salutaire. Ma thèse développe donc un

regard critique à la fois du paradigme dominant et de la critique la plus fréquente de ce paradigme.

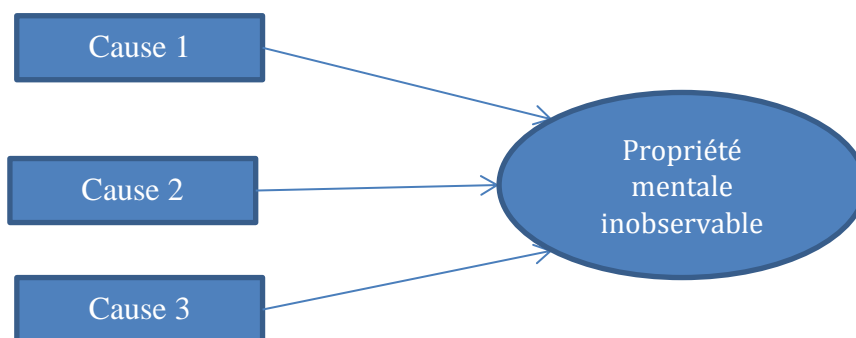
Ma thèse s'appuie sur des réflexions et écrits de nombreux auteurs. Il ne semble donc n'y avoir aucune réelle originalité dans mon développement. Cependant, si je m'autorise à publier cette thèse, c'est que je ne l'ai rencontrée nulle part ailleurs posé de la sorte. Je m'excuse donc par avance si ce que j'écris n'est qu'un plagiat maladroit de ce que d'autres ont déjà formalisé antérieurement. Une telle situation serait bien évidemment frustrante car retirerait toute originalité à mon travail, mais serait en même temps rassurante.

Les deux parties de la thèse

Ce travail discute des variables latentes utilisées en psychologie pour mesurer des propriétés mentales. Mon travail est construit comme une réponse critique à deux discussions différentes portant sur l'usage des variables latentes. La première porte sur ce que Vautier, Veldhuis, Lacot, & Matton (2012) appellent « une crise » au cœur de la démarche scientifique en psychologie (« *The persistence of critical publications about measurement in psychology bears witness to a profound crisis at the heart of the discipline* », p. 810). La discussion sur « la mesure » en psychologie renvoie à des champs très différents que je résume par le terme de *discussion épistémologique* (au sens plutôt anglophone du terme : théorie de la connaissance, sans prétention philosophique au sens plus large du terme, donc un peu en décalage avec le terme *épistémologie* francophone).

Une seconde discussion connexe porte sur un cadre particulier de formalisation des variables latentes : le cadre formatif. Le cadre formatif propose de mesurer une propriété mentale non plus par ses manifestations, mais par ses causes comme le montre la Figure 2.

FIGURE 2 : MODÈLE FORMATIF



Hardin & Marcoulides (2011) appelaient à une clarification sur ces modèles formatifs conséquents des avis diamétralement opposés qui émergent dans la littérature académique sur la fiabilité de tels modèles de mesure (« *Unfortunately, the result of this body of research is a series of disjointed and contradictory messages that are not only confusing to consumers of this research but also could threaten the advancement of knowledge [...] This cumulative body of research delivers a disturbingly divergent set of messages.* », p. 754-756).

Il peut paraître surprenant de voir ces deux sujets abordés dans une seule thèse. La raison première est qu'à l'origine de cette thèse je me sentais confronté à ces deux problèmes longuement discutés dans la littérature académique sur les variables latentes : 1/ A-t-on le droit d'utiliser des modèles avec variables latentes ? ; 2/ Si oui, a-t-on le droit d'utiliser des modèles formatifs ? J'ai donc mené cette double discussion parallèlement durant ma thèse. Mais d'autres raisons m'ont mené à considérer important d'introduire la seconde partie suite à la première, je les exposerai en préambule de la seconde partie.

La forme de la thèse

Une démarche possible aurait été de discuter, pour une telle problématique, des différents positionnements, pointant leurs apports et limites. Telle n'est pas la démarche suivie par ce travail car il est hors de ma compétence de prétendre couvrir le spectre des positionnements vis-à-vis des variables latentes. Je discuterai bien évidemment de différentes positions académiques, mais mon travail essaye d'apporter des réponses à un questionnement important en psychologie (au regard du nombre d'articles publiés sur cette problématique). J'adopte le parti pris que le lecteur cherche dans ce texte un point de vue particulier. En conséquence, ce texte se veut clairement rédigé dans une démarche discursive, au sens d'un développement particulier inscrit dans une discussion plus large. La difficulté est que j'aborde différentes questions inter-reliées, difficilement développables dans une structure linéaire comme l'est par nature un texte. De ce fait je poserai parfois des prémisses que je ne développerai que par la suite. De plus je m'excuse du style usant parfois de façon récurrente des mêmes arguments, mais défendre une thèse c'est enfoncer un clou. Donc mille excuses au lecteur si je frappe trop sur ce clou.

Par ailleurs, le « nous » est usuellement utilisé dans la rédaction d'une thèse. Cependant, le « nous » intègre le lecteur à la pensée du rédacteur, rendant implicitement le lecteur solidaire

de l'écrit. Comme je le considère, ce manuscrit est à comprendre comme un moment d'une discussion. Son titre porte le terme « *réflexion* » car il s'agit bien d'une réflexion personnelle. Mon choix sémantique est d'inscrire la lecture de ce texte dans une telle discussion et réflexion. Cette thèse ne pose que mes propres convictions. Il me paraît dès lors plus pertinent d'user préférentiellement du « je » plutôt que du « nous » lorsque je développerai mes arguments afin de permettre au lecteur de conserver une distanciation souhaitable pour la discussion critique. Par contre j'userai du « nous » lorsque le développement n'appuiera pas un argument personnel.

Enfin, toute thèse s'appuie sur une lecture critique. Le champ de mon travail est vaste, et l'évolution de l'accès aux bases de données rend ce travail délicat : d'une part la floraison des revues et des articles rend quasiment impossible une lecture exhaustive sur un champ même étroit ; de plus, l'accès aux revues devient complexe, ainsi les bases de données des universités Paris Descartes et Paris Sud et celles de l'INSERM ne me donnent plus accès à certaines revues fondamentales pour mon travail. Néanmoins, je discute d'un certain nombre de travaux publiés et je donne assez fréquemment des parties rédigées d'autres auteurs sans me contenter de ne mettre que la référence. En effet, dans la forme actuelle académique, le renvoi à des articles sert de caution au propos. Or il s'avère parfois (souvent ?) que l'article, qui sert de support à un propos, développe en réalité des aspects contradictoires avec le propos auquel il sert d'argument. Par transparence et pour faciliter un regard critique, j'ai donc décidé d'alourdir ce texte par l'ajout de fragments de textes. Comme ponctuellement la même référence m'est parue utile à des moments différents de mon propos, je retranscris parfois les mêmes propos dans deux sections différentes. J'ai de même décidé de laisser le plus souvent ces fragments de textes dans leur langue d'origine car, n'étant pas traducteur professionnel, mes choix de traduction auraient pu déformer la pensée des auteurs que je cite. Cependant, comme certains textes ont été traduits en français, je me sers dès que possible de ces traductions.

Une discussion

Ce texte est finalisé car il faut bien à un moment donné mettre un point final à une réflexion écrite. Mais ce texte s'est construit dans une discussion continue sur 3 années avec mes deux directeurs de thèse, certains collègues, des ami.e.s, et surtout les rédacteur.e.s de tous les livres et articles que j'ai lus. Ma pensée a énormément évolué durant ce travail pour essayer

de concilier cette tension permanente que nous retrouvons dans toute la littérature académique : rien ne semble justifier, au sens de la science, l'usage des modèles avec variables latentes ; et pourtant ces modèles semblent opérationnalisables. J'ai donc essayé de trouver des réponses aux différents nœuds que la littérature mettait en avant.

Mes interlocuteurs privilégiés furent mes deux directeurs de thèse. A aucun moment B. Falissard ou J. Juhel ont essayé d'orienter mon travail, au sens d'essayer d'influencer mon cheminement. Leur apport a été l'interaction critique nécessaire, avec beaucoup d'humilité de leur part et toujours beaucoup de pertinence. Ce n'est qu'au cours de ce travail qu'a émergé petit à petit ma conviction qu'il fallait décaler le cadre épistémologique usuel que la discussion académique pose le plus souvent sans le questionner. Je sais qu'aucun de mes deux directeurs de thèse ne partage l'entièreté de mon positionnement.

Il y a donc dans la construction de cette thèse une discussion critique. La critique que je pose ne se veut à aucun moment polémique. La critique est toujours critique d'écrits passés, mais elle bénéficie de travaux souvent plus récents que les travaux incriminés. Il faut donc prendre la critique comme un moyen de développer une thèse et non comme un procès envers des écrits plus anciens. Par exemple, tout mon développement pourrait apparaître comme une critique fondamentale d'un des rares livres en français sur la psychométrie (Dickes, Tournais, Flieller, & Kop, 1994). Mais entre ma thèse et ce livre, il y a 20 ans de discussions. On ne peut reprocher aux auteurs de ce livre de ne pas avoir anticipé toute une discussion postérieure à leur livre. Sans la lecture de tous les travaux que je cite, et souvent que je critique, je n'aurais jamais réussi à poser ma proposition de l'usage des modèles avec variables latentes. Ce sont ces travaux antérieurs au mien qui ont été la source de mon travail... et en conséquence ce travail doit lui aussi être lu avec un regard critique, comme interaction nécessaire dans la démarche scientifique car ma thèse ne bénéficiera pas des futurs développements qui nourriront de futures critiques de celle-ci.

L'objectif de ce texte n'est donc pas d'être considéré comme une thèse scellée dans du marbre et validée par mes deux directeurs de thèse, mais bien comme un moment de cette discussion collective. Je reprends à mon compte les mots de Corcuff (2002) « *on doit pouvoir s'émanciper d'une herméneutique par trop positiviste, qui considèrerait le texte comme clos sur lui-même, comme un donné, dont il faudrait le plus scrupuleusement possible enregistrer "le sens vrai" déposé par l'auteur dès le départ. Je me calerai donc sur une herméneutique plus ouverte et pragmatique, pour laquelle le texte est considéré tout au plus comme un ensemble de potentialités, susceptibles d'être actualisées dans l'interaction avec un lecteur situé dans*

un contexte socio-historique différent de l'auteur, et donc porteur de questions spécifiques par rapport à cet auteur » (p. 10).

Structure de la thèse

Ce travail est décomposé en deux parties très distinctes. Si ces deux parties peuvent être lues comme deux blocks séparés, leur connexion est assez évidente : apporter une réponse à la question « *a-t-on le droit de faire ce que nous faisons avec les variables latentes ?* ». Les deux parties seront construites autour d'outils très différents. La première sera une volonté de justifier le cadre épistémologique que je considère nécessaire pour user de mesures en psychologie, et en conséquence je défendrai la thèse que la mesure (par un modèle avec variables latentes) d'une propriété mentale est non seulement scientifiquement valide, mais scientifiquement nécessaire, pour autant que l'on affirme un cadre épistémologique pragmatiste réaliste. La seconde partie usera de techniques usuelles en statistique pour vérifier l'efficacité de modèles formatifs : des simulations Monte Carlo. Elle renverra à des aspects plus techniques, mais se nourrira de la première partie.

Concrètement mon objectif est de repositionner la discussion sur les modèles avec variables latentes. Ma thèse reprend donc pas à pas les problèmes essentiels posés par les modèles avec variables latentes pour amorcer des réponses. Les modèles avec variables latentes doivent considérer comme « réelles » les propriétés mentales auxquelles elles se réfèrent. Cela me permettra de discuter de la nature ontologique des propriétés mentales associées aux variables latentes. Une position réaliste renverra au concept « d'émergence », mais en clarifiant cette émergence. De cette position réaliste, je questionnerai à la fois les concepts inhérents à ces modèles ; et d'autre part la démarche quantitativiste (mesure, validité de ces mesures). Je chercherai à montrer que les propriétés mentales conceptualisées en psychologie nécessitent une clarification et une adaptation des concepts à ce champ. De même, je poserai que la validité de la quantification d'une propriété mentale ne peut être démontrée, sans pour autant nous interdire de le faire ... là aussi en adaptant le cadre formel de la mesure à la psychologie. Tout ce parcours critique me laisse considérer que les modèles avec variables latentes peuvent donner un cadre de modélisation d'une propriété mentale non observable. Je ferai néanmoins un retour critique de l'usage des modèles avec variables latentes.

A partir de cette clarification épistémologique, dans une seconde partie, je discuterai de variables latentes particulières, les construits mesurés formativement. Si jusqu'en 2003 seules

les variables réflexives semblaient légitimes (Borsboom et al., 2003; Jarvis, Mackenzie, Podsakoff, Mick, & Bearden, 2003), les variables formatives ont aujourd'hui une reconnaissance formelle. Je synthétiserai autant que possible la discussion sur ces variables mesurées formativement, puis je rappellerai les menaces méthodologiques liées à ces modèles formatifs. J'apporterai des réponses empiriques, par simulations Monte Carlo, pour apporter un cadre pratique efficient aux praticiens devant utiliser de telles mesures formatives... tout en étant critique de cette formalisation. J'expliquerai donc dans le préambule de la seconde partie les raisons de publier ces travaux de simulations alors que la première partie implique que de tels modèles formatifs sont inopérants. Cette seconde partie est le fruit de travaux menés avec un collègue, Mouloud Tensaout. Je m'attribue dans ce texte les résultats de ces travaux afin que Mouloud puisse conserver son libre arbitre sur mes conclusions. Si c'est moi qui ai codé toutes les simulations et analysé tous les résultats empiriques, ces dernières ont été générées suite à notre collaboration et je n'ai aucune paternité propre sur l'ensemble des résultats développés dans cette seconde partie. Mais j'informe que Mouloud est au courant de ce choix et m'a donné son accord.

La réflexivité du rédacteur

Toute discussion sur la validité « scientifique » d'une thèse ne peut être faite de façon approfondie qu'à partir du positionnement de celui qui pose cette thèse. Bourdieu (2001) appelait le chercheur à avoir une démarche réflexive et à se positionner pour permettre au lecteur d'appréhender sa démarche d'une façon plus critique. Je crois avoir intégré, et sinon l'apport critique de mes directeurs de thèse m'y a contraint, l'*habitus*⁴ usuel d'un chercheur tel que défini par Bourdieu : « *Un savant est un champ scientifique fait homme, dont les structures cognitives sont homologues de la structure du champ et, de ce fait, constamment ajustées aux attentes inscrites dans le champ. Ces règles et ces régularités, qui déterminent, si l'on veut, le comportement du savant, n'existent en tant que telles, c'est-à-dire en tant qu'instances efficientes, capables d'orienter la pratique des savants dans le sens de la conformité aux exigences de scientificité, que parce qu'elles sont perçues par des savants dotés de l'habitus qui les rend capables de les percevoir et de les apprécier, et à la fois disposés et aptes à les mettre en œuvre [...] Pour résumer, le véritable principe des pratiques scientifiques est un système de dispositions génératrices, pour une grande part inconscientes,*

⁴ « *dispositions acquises... des manières d'être permanentes, durables* » (Bourdieu, 1997) p.22

transposables, qui tendent à se généraliser ». Mais, au-delà d'un cadre social et intellectuel propre au champ auquel appartient tout enseignant chercheur, je crois devoir clarifier mon positionnement philosophique en introduction.

Discutant de propriétés mentales, je discuterai des variables latentes à partir d'un positionnement clairement matérialiste (et donc athée). Le courant « matérialiste » couvre de fait des démarches et positionnement très variés, et je me devrai de clarifier mon positionnement au cours de cette discussion. Si je pose une thèse, l'histoire de la science montre que la « vérité » que recherche la science n'est jamais acquise, toute théorie n'est vraie que relativement à nos connaissances du moment. Depuis Kuhn, l'évolution de la science n'est parfois plus vue comme une évolution par « accumulation » mais par « révolutions ». Le débat reste ouvert sur la tension « accumulation » vs « révolution », mais ceci impose une humilité au scientifique. Toute théorie est inscrite dans un contexte idéologique au sens large, c'est-à-dire dépendant du cadre social et historique, dépendant des courants scientifiques dominants de son époque, dépendant de la critique existante, ... Il y a donc toujours une forme de relativisme associée à une théorie. De plus, toute théorie, toute discussion, se fait par le prisme du langage qui lui aussi est un élément dissonant dans la démarche scientifique car encapsulé dans une époque.

La science a conscience de ce relativisme, allant jusqu'au post-modernisme pour ne voir dans le discours scientifique qu'un « discours » (Lyotard, 1979). Tel n'est pas ma position. Si le scientifique est limité par un carcan culturel, historique et linguistique ; si le scientifique est limité par ses capacités cognitives ; la démarche scientifique s'inscrit toujours dans une réalité matérielle. C'est la réalité matérielle, dont le discours fait partie, qui est toujours à la base de la démarche scientifique. Tout concept, toute théorie, a une finitude, mais les évolutions/révolutions scientifiques se font toujours par une confrontation avec notre perception de la réalité matérielle. Si les concepts et théories évoluent, si notre regard se modifie sur le monde (la vue elle aussi n'est jamais culturellement neutre), la réalité reste l'élément premier de notre interrogation scientifique, c'est elle qui interpelle le scientifique. C'est la réalité matérielle (le discours en faisant partie) qui agit et pousse le scientifique à critiquer, à théoriser, à conceptualiser, pour apporter un cadre interprétatif et explicatif de cette réalité, même si ce cadre formalisé reste in fine relatif. Comme le dit Bourdieu (2001) : « *les chercheurs, dans leur confrontation, acceptent tacitement l'arbitrage du réel* » (p 137) Mon positionnement s'affirme donc clairement matérialiste et réaliste, mais sans contradiction avec une forme de relativisme et donc de constructivisme. Je reviendrai sur toutes ces questions dans ma thèse. Mais fondamentalement je peux faire mien ce cadre

matérialiste posé par Marx (1859/1972) « *Le concret est concret parce qu'il est la synthèse de multiple déterminations, donc unité de la diversité. C'est pourquoi il apparaît dans la pensée comme procès de synthèse, comme résultat, non comme point de départ, bien qu'il soit le véritable point de départ et par la suite également le point de départ de la vue immédiate et de la représentation [...] La méthode qui consiste à s'élever de l'abstrait au concret n'est pour la pensée que la manière de s'approprier le concret, de le reproduire sous la forme de concret pensé. Mais ce n'est nullement là le procès de la genèse du concret lui-même [...] Pour la conscience le mouvement des catégories apparaît comme l'acte de production réel – qui reçoit une simple impulsion du dehors et on le regrette – dont le résultat est le monde ; et ceci (mais c'est encore là une tautologie) est exact dans la mesure où la totalité concrète en tant que totalité pensée, en tant que représentation mentale du concret, est en fait un produit de la pensée, de la conception ; il n'est par contre nullement le produit du concept qui s'engendrerait lui-même, qui penserait en dehors et au-dessus de la vue immédiate et de la représentation, mais un produit de l'élaboration de concepts à partir de la vue immédiate et de la représentation [...] Après comme avant, le sujet réel subsiste dans son indépendance en dehors de l'esprit [...] Par conséquent, dans l'emploi de la méthode théorique aussi, il faut que le sujet, la société, reste constamment présent à l'esprit comme donnée première » (p. 165-166)*

Ma critique du cadre épistémologique, le plus souvent implicite, en psychologie (l'empirisme-réaliste) ; ma conscience de l'inscription sociohistorique de tout discours, ne m'entraîne donc pas pour autant à glisser dans une sorte de dadaïsme⁵ épistémologique au sens de Feyerabend (1975/1988), c'est-à-dire sans suivre Feyerabend qui appelait à « *saper l'autorité de la Raison* » (p. 30)... même si je relativiserai « la raison » en psychologie par la suite pour l'inscrire dans une démarche pragmatique (au sens du courant souvent appelé « pragmatisme américain ») et réaliste. La discussion comparative de Bourdieu (2001) sur le champ artistique vs le champ scientifique pointe que le champ scientifique a pour enjeu la représentation du réel et « *accepte tacitement l'arbitrage du réel* ». Je reviendrai sur ce réalisme en psychologie dans le texte, mais c'est cet arbitrage du réel qui autorise à différencier l'Art de la Science

⁵ Je me permets une clarification sémantique : si on associe à Feyerabend le plus souvent le terme anarchiste, comme le titre de son livre le laisse croire « *Contre la raison, esquisse d'une théorie anarchiste de la connaissance* » ; en accord avec Feyerabend le terme dadaïsme me paraît plus juste qu'anarchisme pour qualifier sa critique (cf Feyerabend, 1975, note 2, p. 18). Feyerabend défend en effet en épistémologie une démarche proche des membres du courant Dada en Art, qui avait pour programme un anti-Art. Comme le dit Feyerabend, le terme anarchiste renvoie à un courant politico-philosophique dont Feyerabend peut difficilement être considéré comme porte-parole (« *Cependant l'anarchisme... a des caractéristiques que je suis loin de défendre... C'est pour cette raison que je préfère maintenant utiliser le terme dadaïsme* » note 2 p. 18).

dans la construction du discours critique... et donc à rejeter le dadaïsme comme cadre épistémologique en science.

En conclusion, ce positionnement philosophique renverra à une démarche pragmatique que je clarifierai dans mon propos. Le pragmatisme est souvent (Dewey, Putnam pour exemple) associé à une démarche « politique » au-delà de la discussion épistémologique, au sens d'une philosophie de la liberté humaine ; et donc critique de l'enfermement intellectuel et social (Dewey se réclamait du courant socialiste par exemple). Ma réflexion est sans aucun doute nourrie de cette volonté critique au sens politique du terme, et je le revendique sans pudeur⁶ car le lecteur devinera aisément que c'est cette volonté critique qui pilote plus ou moins consciemment ma démarche et mes conclusions, et donc en accord avec Gould (1997) : « *Les scientifiques évitent souvent de mentionner des engagements de ce type car, selon le stéréotype le plus répandu au sujet du savant, la rigoureuse impartialité paraît être la condition sine qua non de la recherche de l'objectivité désintéressée. Je considère ce point de vue comme l'une des affirmations les plus erronées, et même les plus nuisibles, de celles qui sont couramment énoncées dans ma profession* » (p. 33).

⁶ Je suis un acteur de longue date de l'action dite « militante » dans des structures syndicales et/ou politiques considérés comme radicalement critiques de l'ordre social actuel.

Première partie : REFLEXION EPISTEMOLOGIQUE SUR LES
PROPRIETES MENTALES ET VARIABLES LATENTES

1. Ontologie, épistémologie et propriétés mentales

Le premier chapitre de cette thèse posera les fondements ontologiques et épistémologiques que je considère devoir être pour travailler sur des propriétés mentales. Je reviendrai sur tous ces points dans le second chapitre pour discuter d'une façon critique de la mesure de propriétés mentales.

1.1. La nature ontologique des propriétés mentales

Les variables latentes sont devenues standards dans la littérature académique en SHS. Plus précisément, en psychologie, une variable latente mesure une propriété mentale inobservable d'une personne.

Peu d'articles discutent de la nature ontologique des propriétés mentales formalisés par des variables latentes, exceptés de rares articles qui renvoient à une caractérisation vague (« *l'énergie* », Van Der Maas et al. (2006) ou « *puissance* », Van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011). Markus & Borsboom (2013a) ont publié un livre que l'on peut considérer de référence au regard de la renommée des auteurs et de l'objet du livre (livre qui revient sur la quasi-totalité des discussions relatives aux variables latentes). Mais la nature des propriétés mentales n'est pas réellement discutée dans ce livre. Pour Markus et Borsboom, ces propriétés mentales sont « réelles » sans que soit explicité ce que l'on entend par « réel ». Leur dialogue à la fin de l'ouvrage montre que ces deux auteurs n'ont pas forcément le même regard sur la réalité d'une propriété mentale. Ils discutent de l'attribut⁷ « *Bachelorhood* » qui est « *male and eligible for marriage* ». Pour Markus (p 309) « *The usual interpretation [...] is to deny that bachelorhood exists as an independent entity* » ; pour Borsboom « *does the attribute of "bachelorhood" exist? My answer would be positive* » (p. 310). Leur exemple est significatif de la difficulté de définir comme réelle une propriété mentale. La nature « réelle » vs « instrumentale » des propriétés mentales associées aux variables latentes reste une question ouverte dans la littérature et rarement discutée, pour exemple : « *Much of the philosophical debate which takes place on latent variable models centres on reification [...] However, the*

⁷ Je reprends le terme « attribut » que ces auteurs utilisent

usefulness and validity of the methods to be described in this book do not depend primarily on whether one adopts a realist or an instrumentalist view of latent variables » (Bartholomew & Knott, 1999, p. 2).

Il me paraît indispensable d'amorcer cette discussion en discutant de la nature ontologique des propriétés mentales. Car à ne pas définir ce que l'on entend par « la réalité » d'une propriété mentale, l'ouvrage de Markus & Borsboom (2013a) (et bien d'autres auteurs) laisse subodorer que des approches plus récentes, basés sur une vision en réseaux et/ou dynamique des processus cognitifs, sont des voies potentiellement plus efficaces. La raison souvent évoquée est que ces autres méthodes seraient plus proches de la « réalité » que les approches classiques avec variables latentes. Ce livre de Markus et Borsboom laisse donc en suspens la légitimité des modèles avec variables latentes qui ne pourraient être considérés in fine que comme des instruments parfois utiles mais moins efficaces que d'autres modélisations plus proches de la réalité (Borsboom développe ce point de vue depuis quelques années dans différents articles discutant des *Network Analysis* dont je reparlerai). Cela ne pose aucun problème en soi de récuser des approches utilisées dans un champ scientifique pour proposer de nouvelles approches plus efficaces, au contraire même car toute l'histoire de la science est l'histoire de ces « révolutions » (Kuhn, 1970). Mais, comme j'essaierai de le montrer, les approches alternatives renvoient à d'autres problèmes.

J'ai donc décidé de repartir de l'objet étudié : les propriétés mentales inférées par des variables latentes et des contradictions des approches usuelles.

1.1.1. Le cadre philosophique de la discussion académique

Pour résumer, deux courants dominants existent en psychologie vis-à-vis des propriétés mentales formalisées par des variables latentes (voir Juhel & Gilles (2011)) : 1/ un courant minoritaire (découlant du courant constructiviste ou du courant relativiste) qui considère cette formalisation comme des fictions (construction théorique, heuristique, ...) utiles sans pouvoir considérer que nous mesurons quelque chose de « réel » ; 2/ un courant majoritaire ancré dans une approche empirique-réaliste, pour qui le sens d'inobservable attribué aux propriétés mentales fait référence à une réalité que la variable latente cherche à mesurer.

1.1.1.1 L'empirisme-réaliste

Nous retrouvons appliqué aux variables latentes un vieux débat en philosophie des sciences (Barberousse, Kistler, & Ludwig, 2000; Soler, 2009; Ulises Moulines, 2006). Le courant

positiviste logique a montré ses lacunes à n'utiliser que des données observationnelles. Le courant empiriste-réaliste s'est donc construit en réaction à ces limites. Maraun & Gabriel (2013) rappellent les bases de l'empirisme-réaliste : 1/ La réalité naturelle intègre des phénomènes observables et non observables ; 2/ ces phénomènes existent indépendamment de la capacité perceptive humaine ; 3/ Les concepts scientifiques peuvent être catégorisés en termes théoriques ou observables. Les termes observables sont désignés par les données ; les termes théoriques ne peuvent être désignés sur la base d'observations ; 4/ les termes inobservables sont définis par des lois et hypothèses liées à la théorie (concept ouvert) ; 5/ tant que la théorie est correcte, le terme théorique nomme l'entité causale auquel il se réfère. Pour ce courant la recherche de « la vérité » est inhérente à la science (Soler, 2009), mais l'empirisme-réaliste autorise une démarche plus souple qu'une affirmation de représenter la vérité, considérant que « la vérité » n'est pas atteignable, et que les théories considérées valides sont toujours relatives et questionnables (Slaney, 2001).

Les deux courants (empirisme-logique et réalisme-empirique) basent la science sur l'empirisme, la rationalité, l'objectivité pour décrire et expliquer ou prédire le monde. La science généralement différencie les objets perceptuels (phénomène) et les objets non-perceptuels. Le problème est que les objets non-perceptuels ne peuvent être définis sur la base d'observations, donc sont par nature indéterminés. C'est le problème des objets inobservables discutés en science. Le positivisme logique et l'empirisme-réaliste divergent sur ce point. Pour le positivisme, seuls les objets perceptuels sont scientifiques car il faut une règle de correspondance entre un objet et des données observationnelles pour que l'objet soit interprétable (Moulines 2006 ; Soler, 2009 ; Barberousse, Kistler et Ludwig, 2011). L'empirisme-réaliste a émergé suite à la carence du positivisme logique entre autre sur les objets inobservables. Ce courant admet des objets non-liés aux données observationnelles (termes théoriques) car leur référent est au-delà de l'observation directe. Ainsi, à la différence du positivisme logique qui ne considère que les données observationnelles, l'empirisme-réaliste considère des termes théoriques comme pouvant représenter des objets « réels », dont les données observationnelles sont des manifestations et dont le sens dépasse les données observationnelles. Le courant empiriste-réaliste (au sens large) est devenu le courant dominant en science, et de ce fait la plupart des chercheurs travaillant avec des variables latentes se positionnent plus ou moins consciemment dans ce courant.

1.1.1.2 Le positionnement « réaliste » nécessaire

L'utilisation de modèles avec variables latentes s'inscrit dans une démarche qui considère comme « réelle » la propriété mentale inférée par une variable latente puisque le modèle fait dépendre les variables manifestes (les observations) de la variable latente. Une position « constructiviste » ne peut partir d'une telle formalisation car la variable latente ne se réfère plus à un objet réel, mais devient une variable synthétique construite (« index ») ; en conséquence la modélisation doit s'inscrire sur des modèles qui n'usent pas de variables latentes : Analyses en Composante Principales par exemple.

Les deux schémas ci-dessous illustrent la différence entre une approche dite « réaliste » et une approche « constructiviste/instrumentaliste ». Dans la Figure 4, je symbolise l'index par un carré en pointillé. Je reviendrai ultérieurement sur des phénomènes latents inférés de façon formative, ces derniers se différencient théoriquement du cadre constructiviste (la Figure 2 est donc à considérer comme différente de la Figure 4).

FIGURE 3 : Cadre d'une relation « réaliste » entre un phénomène latent et ses manifestations visibles

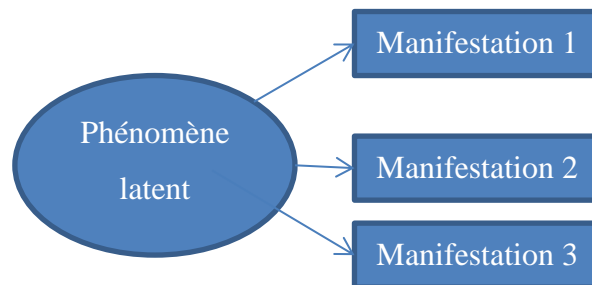
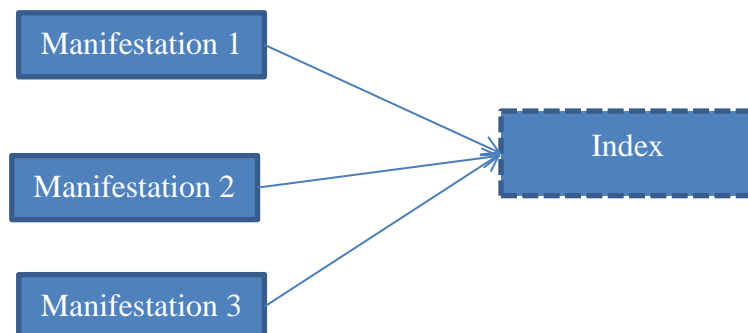


FIGURE 4 : Cadre d'une relation « constructiviste/instrumentaliste » entre des indicateurs et un phénomène



Donc, en accord avec le travail de Borsboom (Borsboom, 2008b; Borsboom et al., 2003, 2004), une position non-réaliste paraît peu défendable dans l'usage des variables latentes modélisé par la Figure 4. Mais encore faut-il s'entendre sur le terme « réaliste », j'y reviendrai par la suite. Pour le moment, je ne discute que de la position « réaliste » vs « non-réaliste » des propriétés mentales associées aux variables latentes.

Je résume les arguments développés par Borsboom et al. (2003, 2004) sur deux axes :

1/ Le rapport à la théorie : si une variable latente ne se réfère pas à une propriété réelle, on ne peut jamais vérifier la validité de la théorie (au sens poppérien). Concrètement, la position non-réaliste interdit de considérer qu'il y ait une hypothèse à tester. Le point de vue réaliste est donc le seul qui permet de tester des hypothèses théoriques. Borsboom et al. (2003) citent Messick (1989) : « *One must be an ontological realist in order to be an epistemological fallibilist* » (p. 26). Cette affirmation de pouvoir valider des théories hypothétiques est contesté par des auteurs à cause des erreurs introduites dans les modèles (comme S. Vautier, voir son site dédié aux modèles psychométriques : <http://epistemo.hypotheses.org/>), je reviendrai sur ce point ultérieurement.

2/ Le rapport aux données : Si nous ne pouvons aller au-delà des données observationnelles, c'est-à-dire au-delà des réponses à des items, les variables estimées se résument à une réduction de données recueillies. On recherchera alors simplement la meilleure synthèse empirique des données, et il n'y a alors plus besoin de variables latentes au sens théorique, les facteurs principaux par exemple suffiront. En conséquence, cela revient à considérer que le construit⁸ équivaudrait à la mesure empirique, et donc différents jeux de données pour un même modèle théorique donneront différents construits. A l'inverse le modèle réaliste considérera différents jeux de données comme émanant d'un seul et même construit. Enfin, valider un construit uniquement sur la base de corrélations, sans référence à une théorie en amont, revient à considérer que deux objets différents mais non différenciables (très corrélés) sont identiques, ce qui est absurde (Borsboom et al. (2004) illustrent cela par la mesure de la taille par le poids). En psychologie, nous travaillons à partir de corrélations entre des variables latentes qui se réfèrent à des propriétés mentales. Or, énormément de propriétés mentales peuvent corrélérer entre elles (Meehl, 1990) ; on ne peut donc pas définir une propriété mentale uniquement par le réseau auquel elle est liée car d'autres propriétés mentales pourraient être définies de la même manière. Les données en psychologie font toujours un focus sur quelques

⁸ Je reprends ici le terme « construit » de Borsboom car son caractère flou (cf mon développement sur le terme « construit » par la suite) permet d'échapper à la tension « réalité » vs « construction » dans ce paragraphe.

concepts sans intégrer l'ensemble des facteurs psychologiques ; et ce focus a besoin d'un cadre théorique en amont. C'est donc le cadre théorique qui nous permet d'aller au-delà des données empiriques.

Cette discussion entre « réalisme » et « constructivisme » est générique à tous les champs scientifiques, mais elle prend une tournure particulière en psychologie lorsque l'on discute des variables latentes. Je reviendrai sur l'ensemble de ces points dans ce travail pour proposer un regard différent.

1.1.1.3 Les problèmes de la position réaliste

Remarquons que Borsboom définit la nécessité de la position réaliste par les contradictions potentielles des positions constructiviste/instrumentaliste avec une démarche « scientifique », mais sans expliciter/justifier par un cadre formel cette réalité des propriétés mentales. Donc cela ne donne à mes yeux aucune légitimité à la position réaliste tant qu'elle n'a pas caractérisé ce qu'elle entendait par « réalisme » de propriétés mentales.

C'est justement parce que les variables latentes renvoient au réalisme que Maraun (Maraun, 2010; Maraun & Halpin, 2008) les critique et considère ces modèles comme une « hérésie » (Maraun, 2012). Les arguments de Maraun posent je pense de vraies questions. Pour Maraun, les concepts sont constituants du langage, dont les règles d'utilisation sont fixées par la linguistique. Les variables sont construites par les scientifiques pour comprendre les phénomènes, les variables sont liées aux données que le scientifique utilise, les variables ne sont donc pas constituant d'une réalité. Les concepts utilisés n'ont pas de réalité matérielle, on n'a donc pas une « variable latente » inobservable, mais on vérifie juste une structure de corrélation des variables manifestes. Il considère que nous ne pouvons pas aller au-delà des données, Maraun (2010) : « *Component models do not "go beyond the data."* » (chapitre III p. 8) ; « *Mere representation is not measurement* » (dernière phrase du chapitre IX). Pour lui : « *the latent variable modeler cannot, and does not, state what kind of thing he is trying to detect (is it, e.g., a material entity or a force?), nor the medium in which it resides (e.g., air, water, blood plasma, atoms of some element, brain tissue), nor the spatio/temporal location at which it is likely to be found (but, then, perhaps he believes that a workable answer is that it is located in an unobservable "domain" or "realm"), nor the mechanism by which it is supposed to bring about the properties {c1,c2,...,cr} of the distribution of the Xj.* » (chap. VII p. 7).

Au-delà de la difficulté linguistique pour conceptualiser des états mentaux (sur lequel je reviendrai), Maraun (2010) demande quelle est la nature réelle de l'attribut et demande une définition spatio/temporelle de celui-ci. Slaney et d'autres (Lovasz & Slaney, 2013; Slaney & Racine, 2013a) considèrent de même que l'attribut « inobservable » est par définition potentiellement observable. C'est donc cette réalité qu'il faut caractériser pour pouvoir justifier l'usage de tels modèles. Or, comme le pointe Maraun, la question de la nature physique des attributs psychologiques utilisés avec des variables latentes est souvent évacuée dans les articles. Par exemple Kane, (2013a) cite Lord et Novick « *As Lord and Novick (1968) put it: "Much of psychological theory is based on trait orientation, but nowhere is there any necessary implication that traits exist in any physical or physiological sense. It is sufficient that a person behave as if he was in possession of a certain amount of each of a number of relevant traits and that he behave as if these amounts substantially determined his behavior.* (p. 358) » et pour Kane « *The use of trait language can be harmless if the trait label is used as shorthand for the observable attributes or as an expression of belief that the observed regularities in performance in the target domain are due to some characteristics of the person, even though we do not know much about this characteristic.* » (p. 32). De même Borsboom, Cramer, Kievit, Scholten, & Franic (2009) « *that one has to make reference to the property measured as a causal force that steers the direction of the measurement outcomes. This is no small matter. It requires a very strong assumption about what the world is like, namely that it contains some property that exists independently of the researcher measuring it. . . . It also obliges the researcher to explicate what the property's structure or underlying process is and how this structure or process influences the measurement instrument to result in variations in the measurement outcomes. This seems to be a very daunting task indeed for many psychological properties that researchers claim to measure.* » (p. 148).

Très clairement, si les modèles avec variables latentes exigent un cadre réaliste, il est impératif d'éclaircir la nature ontologique des propriétés mentales que les variables latentes prétendent inférer. Or la littérature semble la plupart du temps dans l'incapacité d'explicitement la nature « réelle » de ces attributs. Pour reprendre les mots de Desrosières (2008b) : « *Un spectre hante le travail du statisticien, celui de la réalité des objets dont il exhibe la mesure* ».

1.1.2 La nature physique des propriétés mentales

La littérature se contente le plus souvent de considérer que les propriétés mentales sont issues de processus complexes. La littérature use souvent du terme « propriété émergente ». Mais le

terme « émergence » semble devenir un artifice pour éviter le débat sur la nature réelle des propriétés mentales car l'émergence n'est pas souvent explicitée, je reviendrai sur ce point dans mon développement. Il y a donc nécessité de clarifier d'un point de vue réaliste l'ontologie des propriétés mentales, et donc « l'émergence » potentielle associée. Cette tentative de clarification amènera comme conclusion que l'existence de propriétés mentales peut être hypothésée⁹ effectivement comme propriétés émergentes, donc sans pouvoir les définir de façon spatio/temporelle précisément, mais en clarifiant la notion d'émergence et quelles sont ses implications.

1.1.2.1 La vision modulaire du cerveau

Pour la théorie de l'identité en philosophie de l'esprit, la mesure psychologique est redondante avec la mesure neuronale car la propriété mentale est ontologiquement réductible à l'état physiologique. Si une propriété mentale est inobservable, c'est par ignorance ou faiblesse méthodologique, mais on pourra la réduire à des réseaux neuronaux et ce n'est alors plus nécessaire de la mesurer par un modèle de mesure (variables latentes) utilisé en psychologie. Même sans être d'accord avec la théorie de l'identité, il paraît évident aujourd'hui, à moins d'être dualiste (au sens cartésien du terme), de considérer qu'une propriété mentale est liée au cerveau. Mais le fait qu'un état mental¹⁰ soit caractérisé par un état cérébral ne veut pas dire que l'on peut ontologiquement réduire l'un à l'autre comme le souligne bien des auteurs (L. F. Barrett, 2009a par exemple).

Différents auteurs ont essayé de connecter l'activité neuronales et les propriétés mentales inférées par des variables latentes. Kievit et al. (2011a, 2012) proposent une démarche méthodologique originale pour associer des mesures d'attributs physiologiques et des mesures d'attributs¹¹ psychologiques. Leur démarche propose de relier des mesures neurobiologiques et des propriétés mentales au travers de modèles d'équations structurelles. Leur article se veut une réponse à la critique du réductionnisme des neurosciences que pointent certains chercheurs. D'autres travaux proposent des modélisations similaires (Charlton et al. (2008), Dietvorst et al. (2009) pour exemples). Kievit et al. (2011a, 2012) regardent les relations entre deux niveaux d'analyse : niveau neuronale et niveau psychologique pour, au-delà des

⁹ J'utilise ce verbe inexistant et pourtant plein de sens, Cavalla & Grossmann (2004) « *faire est un outil grammatical que nous pourrions remplacer par le verbe Hypothésier si celui-ci existait en français (il existe en anglais to hypothesize)* » p. 6.

¹⁰ Comme explicité antérieurement, j'use ici de « état mental » pour caractériser une forme particulière d'une « propriété mentale ».

¹¹ Je conserve ici le terme « attribut » utilisé par les auteurs car il correspond à leur démarche physicaliste.

corrélations, pouvoir relier ces deux champs. Au-delà de problèmes méthodologiques (l'usage de modèles formatifs sur lesquels je reviendrai dans la seconde partie de ce travail), tous ces articles se basent sur une vision modulaire du cerveau. De façon plus ou moins claire, à l'image du travail de Kievit et al. (2011a, 2012), la plupart des travaux, usant de neuro-imagerie pour inférer les propriétés mentales, posent implicitement ces dernières comme des modules cérébraux.

Barberousse et al. (2000) rappellent les bases de la vision modulaire du cerveau : « *un système psychologique est modulaire selon Fodor s'il possède un certain nombre de caractéristiques [...] Les modules sont informationnellement cloisonnés et associés à une architecture neuronale fixe.* » (pp. 248-250). Ce qui justifie cette conceptualisation est la nécessité de l'adaptation du système psychologique par le processus d'évolution : « *s'il existe des fonctions psychologiques analysables comme autant d'adaptation à l'environnement, on doit s'attendre à ce que ces adaptations soient relativement indépendantes les unes des autres* » (p. 249), et donc localisées dans des zones différentes. Une cartographie générale des fonctions cognitives du cerveau fait d'ailleurs l'objet d'une agrégation des données des différentes expériences en neurosciences (<http://www.brainmap.org/> ; Laird, Lancaster, & Fox (2005); Poldrack (2006)), mais cette démarche peut être largement discutée voire contestée (au-delà de l'aspect éthique).

1.1.2.2 Critique de la vision modulaire¹²

La neurologie a montré que les fonctions sensorielles et motrices étaient relativement bien localisées dans le cerveau. Il est évidemment séduisant de prédire que les propriétés mentales le seraient également. Or, pour bien des auteurs, une propriété mentale ne peut être localisée de façon absolue dans le cerveau. Les propriétés mentales peuvent être vues comme résultant de processus cognitifs, eux-mêmes vus comme des « fonctions » mentales. D'après cette approche connexionniste, tout processus cognitif peut faire intervenir, à différents moments, des ensembles neuronaux variés et largement distribués. La littérature abonde ainsi d'articles contradictoires sur la localisation de processus cognitifs « élémentaires » et sur les erreurs et de logique scientifiques en neuro-imagerie et d'interprétation des résultats. Je ne donne qu'un petit échantillon de cette littérature.

¹² Je remercie chaleureusement L. Conty pour la relecture de cette partie (L. Conty est MCU HDR en neurosciences à l'Université Paris 8, membre du laboratoire « Psychopathologie et neuropsychologie » et chercheuse associée à l'UMR INSERM-ENS « Neurosciences cognitives »).

L. F. Barrett (2009a, 2009b) discute à partir de différents exemples de l'approche « localisationniste » des fonctions mentales rentre en contradiction avec bien des résultats empiriques. L. F. Barrett rappelle par exemple que l'on ne peut pas isoler une partie du cerveau pour la mémoire, les émotions ou autre. Pour la peur par exemple, on peut trouver des zones neuronales pour le « frozen¹³ », « le sursaut », « la fuite potentielle ». Ces circuits sont distincts, mais aucun d'eux n'est le circuit réservé à « la peur ». McIntosh (2000), à partir de trois exemples, illustre la plasticité du cerveau, les interactions entre zones cérébrales et la non-spécialisation de certaines zones, qui font que pour lui une zone du cerveau ne peut être considérée comme une localisation fiable de processus cognitif précis. Gureckis, James, & Nosofsky (2011) ont montré aussi que des zones différentes du cerveau, analysées au départ comme des fonctions différentes, ne le sont pas forcément. L. F. Barrett (2009a) discute aussi de zones cérébrales initialement dévolues à des tâches spécifiques qui s'avèrent avoir des fonctions plus complexes et moins spécialisées. Anderson (2010) apporte toute une littérature amenant le même relativisme sur les fonctions particulières de certaines zones du cerveau. Pour McIntosh (2000), même pour des tâches simples, il y a certainement différentes régions actives car au moins les fonctions de l'« attention », de la « perception », et de la « motricité » sont actives. Il faudrait mieux parler de réseaux de régions ou d'agrégation plutôt que de localisation de fonction cognitive. Pour Anderson (2010), lors de protocoles expérimentaux, des zones actives peuvent n'être ni nécessaires ni suffisantes. Elles peuvent s'activer sans être liées à la tâche, des zones actives peuvent ne pas être nécessaires. De plus certaines activités cérébrales ne se caractérisent pas tant par les zones activées que les relations entre certaines zones. Pour Harley (2004), en isolant des zones particulières dédiées a priori à certaines tâches, on fait l'impasse sur l'hypothèse que les processus cognitifs puissent en permanence faire appel à l'ensemble du cerveau, et donc, si tel était le cas, la neuro-imagerie n'apportera jamais aucune information sur la cognition. Poldrack (2008, 2010) considère lui aussi qu'il est impossible d'associer une fonction précise à une zone cérébrale, renvoyant la discussion sur la mémoire de travail qui pourrait être inexistante, ou plutôt ne pas être une fonction mais faisant appel à différentes fonctions cognitives. Cependant, comme le souligne L. F. Barrett (2011), cela n'est pas contradictoire avec le fait que certaines régions cérébrales sont dévolues à certaines fonctions.

La neuro-imagerie mesure classiquement des changements métaboliques engendrés au niveau du cerveau par l'activité de neurones (des millions localisés au même endroit) . Ces mesures

¹³ Difficile de traduire ce terme : avoir froid, se figer,

se font à l'aide de la TEP (*Tomographie par émission de positron*) ou de l'imagerie par résonance magnétique fonctionnelle (IRMf). Concrètement ce sont les flux sanguins, et le niveau d'oxygénation lié qui sont indirectement mesurés par sondage toutes les quelques minutes (TEP) ou quelques secondes (IRMf) aux niveaux de voxels (environ 100 000 en IRMf) qui permettent d'obtenir une image du cerveau en trois dimensions (un voxel est un pixel en 3D). La circulation sanguine étant un phénomène physiologique relativement lent (comparé à l'activité d'un neurone), le signal intègre l'activité cérébrale de plusieurs secondes : la résolution spatiale de ces techniques est donc excellente, mais leur résolution temporelle est faible. Il existe des techniques dites électrophysiologiques (telles que l'électroencéphalographie -EEG) qui possèdent au contraire une excellente résolution temporelle mais une résolution spatiale faible. Dans certains cas cliniques, il est possible d'enregistrer l'activité de neurones isolés ou de petits groupes de neurones. Ces enregistrements intracrâniens ont des résolutions spatiales et temporelles toutes deux excellentes. L'ensemble de ces dernières techniques (dites directes), moins connues du grand public et qui font l'objet de moins de publications, présentent les mêmes limites dans leur interprétation des données et les modèles statistiques employés et je me focaliserai par la suite sur l'IRMf pour illustrer mes propos.

Beaucoup d'articles discutent ainsi des erreurs potentielles associées aux résultats en neuro-imagerie qui expliqueraient la fragilité des résultats ayant permis d'associer une zone cérébrale précise à un processus cognitif précis. Sur le plan du recueil des données, la neuro-imagerie souffrirait de positions largement discutables. Pour Brown (2012), la critique la plus importante concerne l'usage du signal BOLD (*blood-oxygenation level dependent*) mesuré en IRMf. Cette critique concerne la relation établie entre les zones cérébrales révélées actives par l'IRMf et les hypothèses cognitivistes, notamment sur l'unicité d'une zone pour une tâche. Brown parle de la *Blobology*¹⁴, c'est-à-dire de la cartographie du cerveau à partir de blobs d'activation révélés par l'IRMf.

Heeger & Ress (2002) considèrent que l'on recueille au mieux une approximation entre l'activité neuronale, la demande métabolique, le flux sanguin et l'oxygénation. D'après Shallice & Cooper (2011) le signal BOLD acquis en IRMf montre des potentiels sans pouvoir assurer pour autant que les zones visualisées (les blobs) soient primordiales à la fonction étudiée. De plus, Sirotnin & Das (2009) montrent que l'interprétation du signal BOLD (relation

¹⁴ Un *blob* désigne l'activité significative d'un ensemble de voxels (un cluster), représentée graphiquement sous la forme d'une tâche de couleur plus ou moins intense selon la valeur statistique du cluster.

linéaire entre les neurones et le signal hémodynamique) est valide pour les signaux visuels, mais ne permet pas de prédire les autres signaux. Enfin, l'hypothèse de l'homogénéité des zones est largement contestable. L'imagerie neuronale se base généralement sur une hypothèse faible : le lien entre les réponses et les zones activées sont monotoniquement liées au signal BOLD de façon identique pour toutes les personnes. De plus, la corrélation entre les états cognitifs et les états du signal neuronal est une corrélation partielle, et leur synchronisation temporelle est loin d'être parfaite. Le signal enregistré, en TEP ou en IRMf, est un simple signal d'activation qui ne peut en aucun cas être interprété en termes d'excitation ou d'inhibition. Cela a pour conséquence qu'un même pattern d'activation observé dans deux situations différentes peut avoir une signification fonctionnelle différente. Sans oublier aussi les effets résiduels des modifications hémodynamiques qui peuvent durer près d'une minute, les artefacts divers liés aux mouvements du sujet, à sa respiration et à son rythme cardiaque et, enfin, l'hétérogénéité du rapport signal-bruit selon les différentes régions du cerveau.

Au-delà du recueil des données largement discutable en neuro-imagerie, l'analyse des données est aussi sujette à caution. Vul & Kanwisher (2010) expliquent les difficultés d'analyser les données en IRMf (masse de données, recueil des données, ...). Un IRMf à faible résolution donne 81 000 données toutes les 3 secondes, à haute résolution 500 000 mesures toutes les deux secondes. Les données en 4 dimensions (3D+temps) sont corrigées, co-enregistrées, transformées, lissées pour donner une image du cerveau (« *resampled, detrended, normalized, smoothed, trimmed (temporally or spatially)* »). Chaque étape se fait avec des paramètres choisis par l'auteur. La difficulté réside dans le fait qu'avec un tel volume de choix, il est difficile de vérifier si ceux-ci ne biaisent pas les résultats. Bagozzi (2011a, p. 93) note : « *With a host of multiple apparatuses used to collect and process data, one can see that the likelihood of method bias is very real.* ». Une technique usuelle est de poser l'hypothèse que les bruits sont indépendants. Cependant, Vul et Kanwisher pointent que cette hypothèse est peu vraisemblable (cf. appendice 2 de Vul & Kanwisher, 2010), rendant l'erreur de première espèce quasi certaine (validé pour des données simulées sur des purs bruits). « *It is quite possible that a considerable number of relationships reported in this literature are entirely illusory* » (Vul, Harris, Winkielman, & Pashler, 2009, p. 20).

Après la phase de recueil et de préparation des données, l'analyse régresse généralement l'activité de chaque voxel. La littérature académique révèle des erreurs usuelles d'analyse des données en neurosciences à ce niveau. Ces erreurs sont conséquentes de la gestion d'une masse de données estimées importante. Un IRMf comporte généralement des dizaines de

milliers de voxels, c'est à dire des dizaines de milliers de tests qui comportent un risque important de résultats positifs faux (erreur de type 2). Le problème soulevé est le contrôle des comparaisons multiples qui n'est pas toujours effectuée dans les articles académiques. Nichols & Hayasaka (2003); Vindras, Desmurget, & Baraduc (2012); Vul et al. (2009); Vul & Kanwisher (2010) posent ainsi le problème de la comparaison multiple effectuée en neurosciences et de la fiabilité statistiques de nombreux résultats en neuroscience. En suivant le protocole usuel utilisé en neuro-imagerie, Bennett, Baird, Miller, & Wolford (2011) montrent qu'un saumon mort peut présenter des activations significatives lors d'une tâche cognitive effectuée sous IRMf.

Vul et al. (2009) et Vul & Kanwisher (2010) pointent une autre erreur méthodologique très fréquente : le biais de sélection ou « l'erreur non-indépendante ». La zone cérébrale analysée (voxels significativement différents entre deux tâches ou deux populations) est très souvent sélectionnée sur la base de résultats statistiques, puis analysée par rapport à d'autres variables. Ce procédé, dit récursif, aboutit à ce que les résultats ne fassent que confirmer des biais de sélection de voxels. Vul et Kanwisher discutent de beaucoup d'articles contenant cette erreur : « *It is rare to find an issue of a neuroimaging journal in which none of the articles have plotted non-independent data* » (Vul & Kanwisher, 2010) et illustrent par simulations à partir de bruits gaussiens les résultats faussement positifs que l'on peut trouver sur la base de ces procédés récursifs.

1.1.2.3 Les bases d'une vision non-modulaire du cerveau

Pour Tiberghien (2007) l'effet mesuré en neuro-imagerie est attribué à une entité cognitive hypothétique que l'on postule unitaire, c'est-à-dire non-décomposable en d'autres sous-processus cognitifs, de plus bas niveau, ou non-attribuables à des processus cognitifs de plus haut niveau. C'est un pari évidemment à très haut risque et qui peut même être théoriquement et logiquement contesté, car il repose à la fois sur une conception strictement modulaire de la cognition et sur un modèle linéaire du cerveau. Anderson (2010) pose une autre manière d'appréhender les réseaux neuronaux en posant l'hypothèse d'une requalification des neurones en fonction des tâches à accomplir. Cette perspective ne contredit pas elle non plus que certaines zones soient dévolues plus spécifiquement à certaines tâches, mais cela relativise énormément la vision purement modulaire du cerveau. Les réseaux neuronaux non seulement seraient en interactions permanentes, mais de plus auraient des fonctions différentes et évolutives. Cette approche permettrait de comprendre les différences de

localisations trouvées entre populations de cultures différentes (cf. les résultats de Chiao & Cheon, 2010). Shimamura (2010) confirme cette critique de la vision réductionniste, une complète explication de la cognition demande de prendre en considération beaucoup de circuits neuronaux et les interactions entre eux. D'ailleurs pour Fodor (1983) le système modulaire pourrait ne pas s'appliquer aux processus cognitifs de « *hauts niveaux tels que le raisonnement ou la prise de décision* » de la pensée.

La vision modulaire des processus mentaux semble donc largement discutable, autant sur ses fondements théoriques que sur les résultats empiriques.

1.1.3 La théorie des systèmes complexes associée au cerveau

La critique de la vision modulaire du cerveau pour des propriétés de « hauts niveaux » renvoie à une vision de ces propriétés mentales comme résultantes d'un réseau en interaction. Cette approche me semble plus pertinente, mais elle obligera dès lors à discuter de ces propriétés comme des propriétés émergentes, en précisant ce que l'on entend par « émergence ».

1.1.3.1 *Attractor Neural Networks*

Une vision des états mentaux comme résultants d'un réseau en interaction permanente semble plus pertinente que la vision modulaire pour des propriétés mentales de « haut niveaux ». Dans un cerveau, il y a en effet près de 100 milliards de neurones qui forment un réseau d'interactions massivement parallèles et l'ensemble des états possibles de ce réseau dépasse le nombre d'atomes dans l'univers (Falissard, 2008a). Plus un processus cognitif est complexe, et plus il ferait intervenir, dans l'espace et le temps cérébral, un nombre important de sous-systèmes selon une dynamique non linéaire (Tiberghien, 2007). Dans les processus cognitifs, il y a une interaction multiplicative, et les interactions se modifient à chaque action d'un nouveau processus (Van Orden & Paap, 1997). Il y a un problème de boucle et de niveaux de phénomène cognitifs interdépendants (Wallot & Van Orden, 2011). De plus, le processus est environnement-dépendant (interaction avec le contexte et l'environnement, Van Orden, Kloos, & Wallot, 2011). La résolution d'une tâche doit donc être vue comme un processus intégrant dans un système des composantes en interaction et en interdépendance, avec des effets connexes dans le temps (Wallot & Van Orden, 2011; Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012).

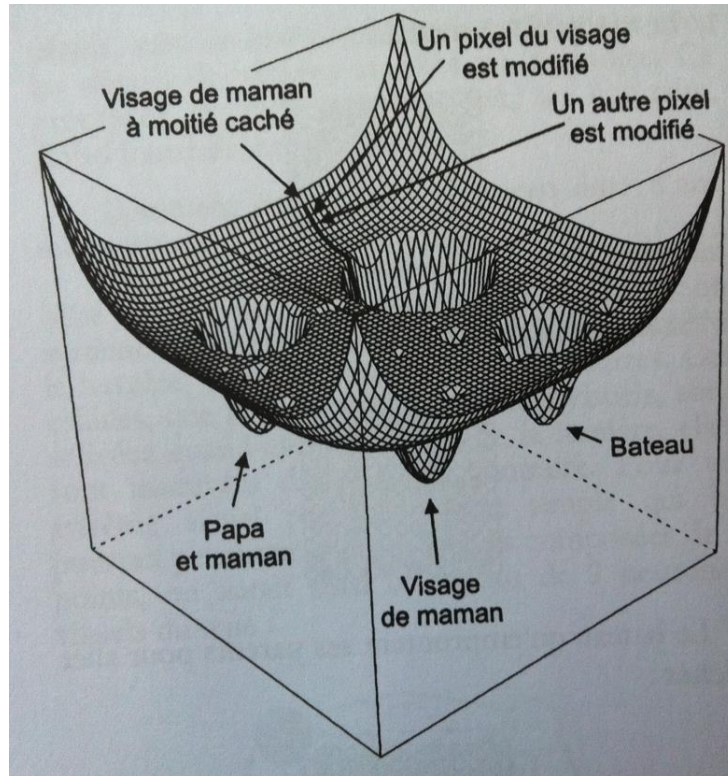
En opposition avec l'approche qui cherche à isoler les processus, des méthodes alternatives partent de l'hypothèse de systèmes dynamiques complexes (phénomènes interagissant). Les

processus cognitifs ne sont alors pas analysés comme des processus moyens avec variations gaussiennes (bruit blanc), mais comme des processus ayant un bruit rose (caractéristique de systèmes dynamiques complexes, dits souvent systèmes chaotiques) : la fréquence d'apparition de l'amplitude d'un processus est inversement proportionnel à sa fréquence ($S(f) = 1/f^\alpha = f^{-\alpha}$) qui se caractérise par une relation linéaire entre le log de la fréquence et le log l'amplitude (le coefficient α étant la pente de cette relation linéaire). Des études récentes montrent que bien des processus cognitifs correspondent au cadre du système dynamique complexe car révélant des bruit roses (Wijnants et al., 2012). Dans de tels systèmes, l'activité de chaque processus est interdépendante des autres et on ne peut analyser la contribution individuelle d'un processus au système, marquant une « coopération » entre les processus (Wijnants et al., 2012). Ce n'est pas une chaîne causale de processus, ajoutant chacun indépendamment une réponse, mais une caractéristique de boucles et de niveaux interdépendants. L'analyse en système complexe semble apporter des réponses en intégrant ce qui est généralement considéré comme du bruit dans l'analyse traditionnelle de la mesure des processus cognitifs (Wijnants et al., 2012). De tels bruits roses sont caractéristiques de systèmes dits « chaotiques » issus de système complexes.

C'est donc sur la base des études des systèmes chaotiques que se sont tournés bien des travaux sur les réseaux de neurones. Ces systèmes chaotiques montrent des dynamiques qui entraînent ces systèmes vers des états stabilisés, les attracteurs du système. Les *Attractor Neural Networks* donnent ainsi un cadre explicatif/illustratif du fonctionnement du cerveau qui semble plus efficient. Pour une discussion illustrée, voir Falissard (2008a). Le réseau de neurones est envisagé comme un système dynamique non-linéaire complexe. Falissard (2008a) précise : « *Le cerveau humain ne fonctionne assurément pas dans la réalité comme ce qui a été décrit [...] Pour décrire et comprendre la nature de la pensée [...] une simplification assez drastique [...] peut s'avérer particulièrement féconde* » (p. 93-95). Une propriété mentale pourrait être alors comprise non comme un ou des modules du cerveau, mais comme un état stabilisé d'un réseau, un attracteur du réseau. Deux personnes ayant des propriétés mentales proches pourraient voir leur réseau être très différent, mais voir celui-ci stabilisé sur des attracteurs « proches » au sens des conséquences propres à ces attracteurs.

Je rapporte une illustration tirée de Falissard (2008a) qui schématise des attracteurs proches pour reconnaître des « objets » (« papa et maman », « maman », « bateau », ...).

FIGURE 5 : REPRESENTATION D'ATTRACTEURS

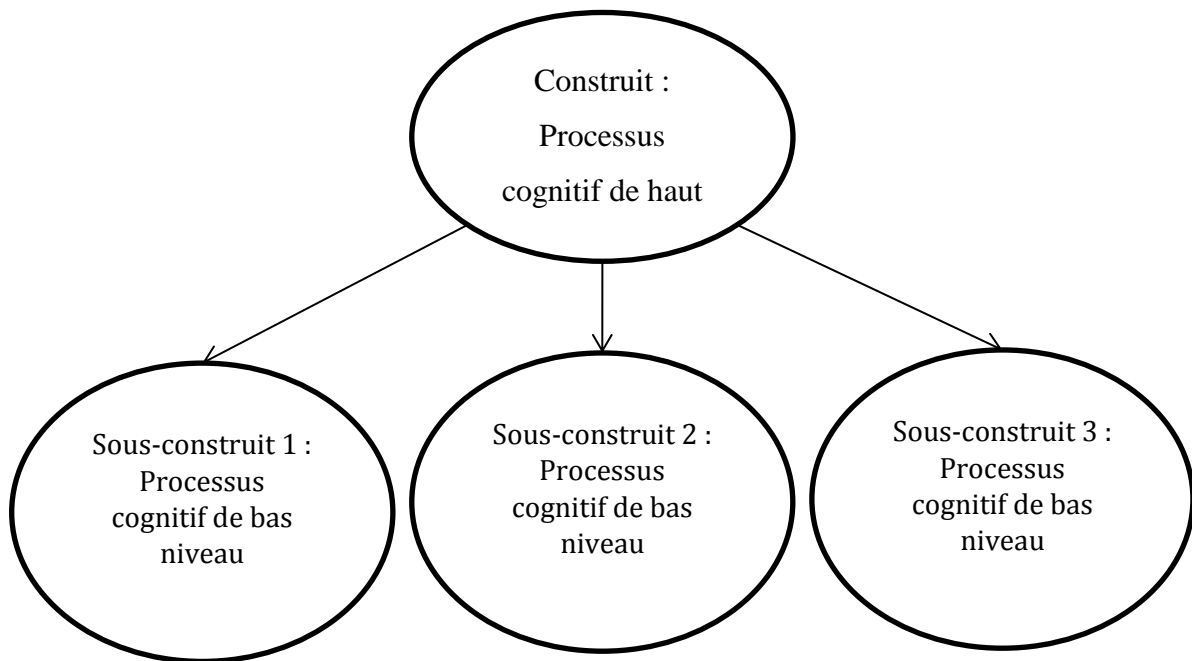


Tiré de Falissard (2008a, p. 18)

1.1.3.2 Variables latente et systèmes dynamiques complexes

Cette vision en système dynamique complexe du cerveau renvoie des questions sur les approches proposées par certains auteurs pour modéliser cette complexité avec des variables latentes. La vision des processus cognitifs comme un système complexe a depuis quelques années été au cœur de discussions sur les variables latentes. La complexité est souvent appréhendée avec les variables latentes par des modèles hiérarchiques de construits et sous-construits (McGrath, 2005a).

FIGURE 6 : SCHEMATISATION D'UN MODELE HIERARCHIQUE AVEC VARIABLES LATENTES



McGrath (2005a) considère que la tendance à conceptualiser les phénomènes psychologiques par des construits hiérarchisés renforce la vision modulaire, et est donc largement critiquable. En effet, les variables latentes modélisées de façon hiérarchique présupposent la concaténation logique des processus en cause dans la tâche (Michell, 1999; Wallot & Van Orden, 2011). Wallot & Van Orden (2011) rappellent par exemple que pour la lecture agissent des sous-composantes : sensation, perception, lecture et articulation qui sont supposées concaténer leurs effets de façon ordonnée dans le temps, chacune affectant la suivante. Il y a une hypothèse d'interaction additive, or la linéarité et l'additivité sont le plus souvent absentes des processus cognitifs. La réalité des processus cognitifs semble bien différente qu'un simple ordonnancement de fonctions cognitives (Wallot & Van Orden, 2011).

Il apparaît donc problématique au regard d'une vision comme « processus complexe » de modéliser par un modèle simplement hiérarchique avec variables latentes les processus cognitifs car les liens causaux linéaires et additifs sont en contradiction avec les interactions permanentes et différenciées. Des auteurs ont essayé de relier cette complexité et les variables latentes. Bartholomew, Deary, & Lawn (2009) proposent un modèle intégrant différents liens neuronaux, mais le modèle impute un score individuel à chaque lien pour construire un score global à un item. Leur modèle se construit en phase avec la neuro-imagerie. Dans celui-ci, l'« énergie » déployée est mesurable par le signal BOLD. La mesure du test deviendrait la mesure de l'énergie global dépensée, plus il y a de liens plus l'énergie

est forte et plus le score est élevé (avec une vision clairement phrénologique puisque dans ce modèle la taille du cerveau est corrélée à l'intelligence). Cette modélisation est là aussi très hiérarchisée et modulaire, et la capacité du réseau n'y est que la somme des capacités individuelles de chaque module, ces dernières étant mesurées par la neuro-imagerie. Nous avons donc de nouveau un chaînage linéaire de fonctions cérébrales en contradiction avec les systèmes dynamiques complexes. De plus, le score individuel de chaque lien n'est pas explicite. En quoi un lien neuronal peut donner un score de façon isolée ? Cette approche paraît fragile au regard de notre critique de la capacité de la neuro-imagerie à capter des processus complexes, et des apports des systèmes dynamiques complexes comme cadre de compréhension des processus cognitifs complexes.

En résumé, les approches traditionnelles essayant d'intégrer le « complexe » dans les modèles avec variables latentes semblent échouer. La prise en compte des processus cognitifs comme phénomènes complexes non-hiérarchisables semblent compromettre les variables latentes telles qu'utilisées dans les tests. Van der Maas et al. (2011) arrivent à ce constat : « *the lack of a psychological justification for this key property of the measurement model compromises test validity. The reason is that validity requires a causal mechanism linking the trait or ability with the item responses (Borsboom & Mellenbergh, 2007; Borsboom et al., 2004). In the absence of such a mechanism, the relation between the targeted attribute and the item responses is essentially a black box, and the psychological appropriateness of the function that describes this relation becomes an article of faith.* » (p. 353).

1.1.3.3 Network Analysis

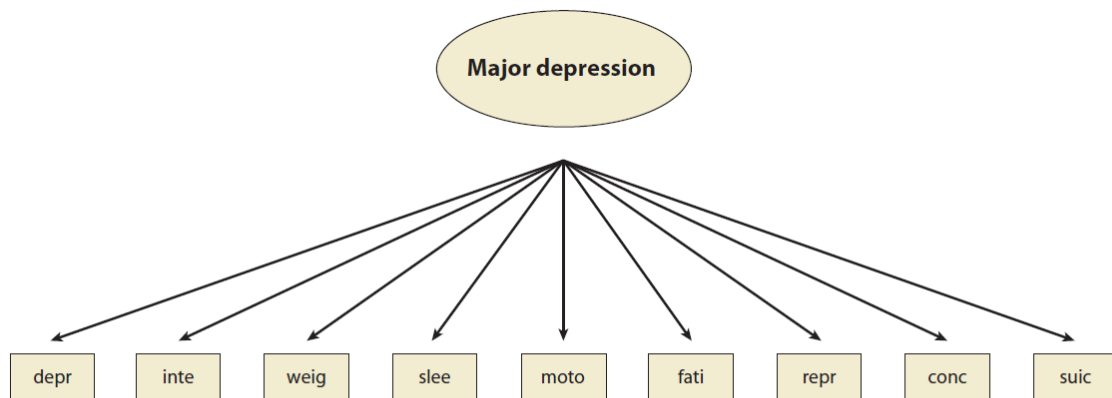
Différents auteurs (souvent liés par les publications, voir pour exemple Borsboom & Cramer (2013); Bringmann et al. (2013); Cramer et al. (2012); Schmittmann et al. (2013)) ont proposé une approche en rupture avec les variables latentes traditionnelles pour appréhender la complexité des propriétés mentales : *Network Analysis*. Je me dois de préciser que le terme de « *network* » ici ne renvoie pas aux *Neural Networks* discutés précédemment. C'est aussi une modélisation en réseau, mais différente. J'utiliserai donc les deux termes envoyant à des modélisations différentes : *Neural Networks* pour la formalisation du système de neurones en réseau avec attracteurs ; et *Network Analysis* le cadre formalisé par Schmittmann, Cramer, Borsboom, discuté ci-dessous.

Cette approche (*Network Analysis*) me paraît intéressante pour essayer d'appréhender la complexité des propriétés mentales ... mais renvoie à d'autres problèmes par ailleurs comme

je vais le discuter. Borsboom & Cramer (2013) développent un cadre illustratif de cette approche sur lequel je vais m'appuyer. Avec le *Network Analysis*, la propriété mentale n'est ici plus discuté comme existante (Borsboom et al., 2003) ou mal posée (Maraun, 2010) ou « *agnostique* » (terme employé dans l'article de Schmittmann et al. (2013) et qui renvoie à Cronbach & Meehl (1955)), mais uniquement appréhendé par un système relationnel entre ses manifestations. Borsboom & Cramer (2013) expliquent (à partir de l'exemple de la dépression, *PD*) : « *in a latent variable model, PD is a latent variable that causes its observable symptoms, such as experiencing panic attacks. From this psychometric perspective, symptoms are regarded as measurements of a disorder, and in accordance, symptoms are aggregated in a total score that reflects a person's stance on that latent variable [...] we argue that complex network approaches, [...] have the potential to provide a way of thinking about disorders that does justice to their complex organization [...] It provides a middle ground in which disorders exist as systems, rather than as entities* »

L'approche par variable latente va modéliser la dépression de la sorte par exemple :

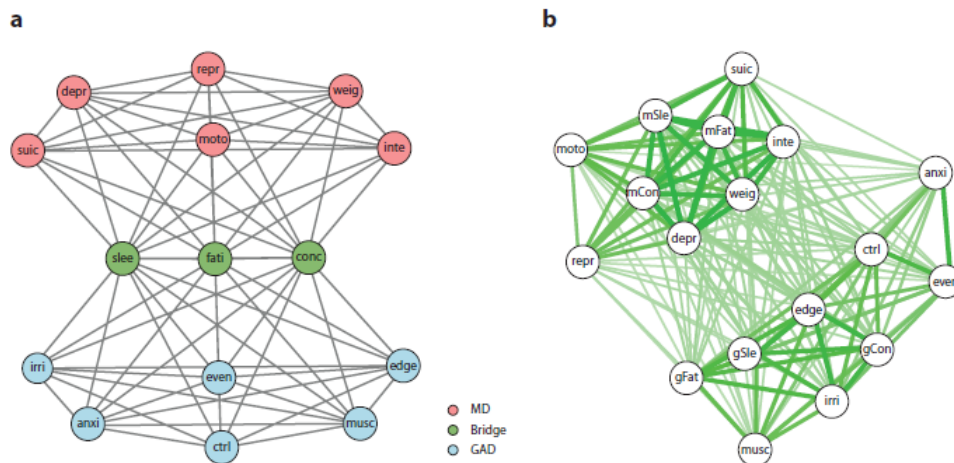
FIGURE 7 : APPROCHE PAR VARIABLES LATENTES DE LA DEPRESSION



Tiré de Borsboom & Cramer (2013, p. 94)

L'approche *Network analysis* va développer un réseau de relations entre les manifestations :

FIGURE 8 : APPROCHE *NETWORK ANALYSIS* DE LA DEPRESSION



Tiré de Borsboom & Cramer (2013, p. 99)

Pour ces auteurs, nous posons en médecine qu'une cause entraîne des effets. Mais une maladie peut ne pas générer de symptômes visibles (cancer par exemple). En psychologie c'est rarement le cas et la propriété mentale est appréhendée par ses manifestations : « *the proposition that mental disorders cannot be separated from their symptoms* ». Un point important pour eux est la fonction de l'effet sur la cause (interrelation entre la propriété mentale et ses manifestations) : « *the relation between symptoms and disorders has to be conceptualized differently, not only because mental disorders are not identifiable as separate disease entities, but also because there appear to be many direct causal relations between symptoms [...] mental disorders may be caused by the direct activation of symptoms through* ». D'où pour ces auteurs la nécessité de partir d'un réseau « *Instead of interpreting symptoms as a function of a set of underlying/latent disorders, the network approach conceptualizes symptoms as mutually interacting, often reciprocally reinforcing, elements of a complex network [...] These symptoms are strongly correlated because they are part of the same system, i.e., because they causally influence one another* ».

Cette approche propose une solution à la difficulté de caractériser des propriétés mentales. Appréhender les manifestations non plus comme de simples conséquences, mais comme un réseau inter-relié ouvre des possibilités d'analyse que les modèles usuels avec variables

latentes semblent interdire. Cependant, je suis critique sur cette approche pour différentes raisons.

Un premier problème à mon sens est la nature de la propriété mentale. Dans cette approche, de façon explicite, la propriété mentale étudiée (la dépression dans l'exemple de Borsboom & Cramer (2013)) disparaît. La propriété mentale est appréhendée par un réseau complexe de manifestations et seules les manifestations sont utilisées (« *disorders exist as systems, rather than as entities* », Borsboom & Cramer (2013)). Cela renvoie à une propriété mentale apparemment non réelle. Car la propriété mentale n'est pas la cause de ces manifestations, elle est en quelque sorte une synthèse abstraite des relations complexes de manifestations : « *As an important corollary, this means that disorders cannot be causes of these symptoms—at least, not in the way that tumors and bacterial infections are causes of symptoms in medicine. This strongly suggests that the treatment of disorders as causes that exist independently of the symptoms used to identify them involves an unwarranted reification (Hyman 2010).* », et « *if MD does not exist as an entity that exists independently of its symptoms* » (p. 97).

Cela peut être une hypothèse de nier l'existence d'une propriété mentale mais, et c'est un problème, en même temps les auteurs de ces modèles considèrent que la propriété mentale existe réellement : « *we do not resort to conventionalism to claim that mental disorders are constructed, like the concept of a yuppie is constructed out of the properties of being young, urban, and financially well off. There is definitely something real about mental disorders, but what?* » (Borsboom & Cramer, 2013, p. 115). Cramer et al. (2012) considèrent « *a 'trait' is a 'common cause', a structure that 'explains' the stable level of functioning around which a certain variability in 'states' revolves [...] the network perspective is thus 'not' contradictory to current trait theories. Networks result in traits too, the only difference with current trait theories being that in the latter case, traits are most typically explained with a latent variable model in mind.* » (p. 412-419-420) et « *Also, we readily acknowledge that some of the hypotheses that follow from the network perspective could in principle be tested with latent variable techniques.* » (p. 429)

Si la propriété mentale n'existe que par le réseau complexe développé, cela demande une discussion conceptuelle pour discuter des raisons de considérer que la propriété mentale n'existe pas en tant que telle ; et que donc la littérature académique a considéré à tort l'existence d'une propriété mentale alors que les « symptômes » de cette propriété mentale ont une sorte d'autonomie et existent comme inter-reliés sans devoir faire appel à cette propriété mentale. Si la propriété mentale existe, elle ne peut être considérée comme

représentée par un réseau complexe de conséquences de cette propriété mentale, même si ce réseau apporte de l'information sur cette propriété mentale. On doit essayer de la comprendre en tant que réalité. Il y a donc une tension chez eux entre l'abandon des propriétés mentales en tant qu'entités réelles (modélisées par variables latentes)... et en retour la nécessité d'affirmer que ces propriétés mentales existent comme des objets réels, formalisables visiblement par des variables latentes.

Un second problème lié est le choix de conserver des manifestations dans leurs exemples qui proviennent d'un cadre conceptuel autre que le leur. En effet, (pour reprendre les exemples utilisés par ces auteurs) les symptômes de la dépression ou les dimensions de traits de personnalités utilisés dans les mesures de ces propriétés mentales ont été développés dans le cadre d'une approche « cause → conséquence ». Si la cause disparaît, si nous devons analyser uniquement un réseau de faits observés sans recours à une propriété mentale causale en amont, pourquoi ne retenir que les manifestations introduites par la théorie des traits. Il faudrait considérer bien d'autres manifestations potentielles, certaines épurées lors de la construction des tests ; d'autres négligées dans un cadre conceptuel différent.

Enfin, l'introduction d'un tel réseau me semble reconstruire une complexité si nous cherchons à intégrer l'ensemble des relations entre les « symptômes » (au sens large) d'une propriété mentale. La modélisation en *Network Analysis* propose différentes possibilités de liaisons entre les manifestations (pas seulement linéaire) et il me semble que très rapidement nous pouvons aboutir à une nouvelle complexité inintelligible d'une part, et essentiellement basée sur des choix du chercheur d'autre part. La complexité ici n'est plus dans la tentative de modéliser un réseau complexe de processus cognitifs en amont générant une propriété mentale, la complexité ici est un réseau de manifestations en aval d'une propriété mentale (qui disparaît du modèle).

En conclusion, bien qu'ayant des potentiels réels l'approche en *Network Analysis* ne résout pas à mes yeux la question de la manière d'appréhender les propriétés mentales comme des objets réels en tant que tel.

1.1.4 La propriété mentale comme phénomène émergent ?

Une réduction des propriétés mentales aux neurones, rentre en contradiction avec toute une littérature qui conteste cette réduction. L. F. Barrett (2009b, 2011) reprend à son compte la conceptualisation des propriétés mentales comme conséquentes d'un réseau complexe, ce qui pour l'auteure renvoie les concepts analysés en neuroscience et analysés en psychologie à

deux niveaux différents. La propriété mentale est alors considérée comme une « propriété émergente » du réseau de neurones. Mais l'émergence est souvent qu'un simple artifice explicatif. Comme le soulignent Varela, Thompson, & Rosch (1999, p. 192) « *Du moins dans le cognitivisme, le réalisme est-il explicite et fait l'objet d'argumentation ; mais dans l'approche de l'émergence, il s'introduit souvent de manière tacite et n'est tout simplement pas questionné* ».

Il faut donc à mes yeux se tourner vers le champ de la *philosophie de l'esprit* qui discute abondamment de ce concept d'émergence des propriétés mentales, notamment avec un débat entre le physicalisme réductionniste et le physicalisme non-réductionniste. Je reprends la définition du physicalisme donnée par Kistler dans son introduction du tome 1 du livre de J. Kim (2009), c'est-à-dire une définition en rupture avec le dualisme : « *tous les objets sont physiques ou exclusivement composés d'objets physiques ; et toutes leurs propriétés sont physiques ou réductibles sans reste à des propriétés physiques* » (p. 10)

1.1.4.1 Le physicalisme non-réductionniste

Le physicalisme réductionniste « *soutient que les explications de haut niveau (psychologiques, par exemple) sont réductibles à des explications en termes physiques, ce qui implique la plupart du temps que les entités auxquelles font référence les explications psychologiques sont entièrement descriptibles en termes physiques.* » (Van de Vijver, 1997, p. 3). Pour le physicalisme non-réductionniste : « *il n'existe rien d'autre dans le monde que des entités décrites par la physique fondamentale, et les agrégats formés à partir de ces entités. Cependant, l'agrégation des entités fondamentales conduit à partir d'un certain niveau de complexité, à des totalités gouvernées par des lois d'un niveau différent de celui de la physique fondamentale* » (Barberousse et al., 2000, p. 221). Ces entités fondamentales sont dites « émergentes », impliquant différents niveaux de réalité. Il y a donc une vision non-unificatrice de la science associée à une vision physicaliste de celle-ci. Comme le note Kistler (2007) : « *La réduction, et les affirmations de réductibilité ou d'irréductibilité ne concernent que les théories qui rendent compte des phénomènes et non ces phénomènes eux-mêmes. Qu'un concept soit réductible ou non, réductionnistes et émergentistes s'accordent sur la thèse ontologique du physicalisme (minimal) selon lequel tous les phénomènes naturels sont objectivement déterminés par les états de choses physiques* ».

L'émergence est souvent associée au concept de « survenance » (*supervinience*) d'un ensemble de propriétés sur un autre ensemble de propriétés. C'est un concept permettant une

vision non-réductionniste du matérialisme. Van de Vijver (1997): « *On veut donc que le mental soit autonome, mais en même temps on ne veut pas tomber dans le piège de l'anti-matérialisme et considérer, comme certaines formes de vitalisme ou de dualisme, qu'il faut toujours un supplément "antimatériel" pour obtenir de vrais états mentaux.* » (p. 6) et il précise (p.6 et 7) 1/ « *il n'y a pas d'états mentaux à côté des états physiques [...] chaque entité mentale est constituée par des entités micro-physiques et chaque régularité est gouvernée par des régularités micro-physiques* » ; 2/ « *les propriétés mentales ne sont pas identiques, ni réductibles aux propriétés et relations physiques* » ; 3/ « *chaque propriété mentale est réalisée par des propriétés physiques appropriées, ou survient sur ces propriétés, ce qui veut dire que même si chaque état mental est constitué d'états micro-physiques, et même si chaque régularité mentale est gouvernée par des régularités micro-physiques, il n'est pas pour autant possible de décrire ou d'expliquer de manière significative les propriétés mentales au niveau supérieur en termes micro-physiques (autonomie du mental)* » ; 4/ « *il y a une forme de réalisme mental qui considère que les propriétés mentales sont causalement efficaces : le mental a des pouvoirs causaux propres, différents des pouvoirs causaux des propriétés physiques ou biologiques, et nécessite ainsi des explications causales propres.* ».

La théorie de la survenance permet de lier les deux niveaux (micro, au niveau des neurones ; macro, au niveau des propriétés mentales) en faisant le pont entre les apports des neurosciences et les concepts utilisés en psychologie sans tomber dans un réductionnisme biologique étroit. La survenance, qui prolonge le concept « d'émergence », discute des fonctions cognitives complexes et de leurs relations causales qui ne peuvent se réduire aux éléments élémentaires (les neurones activés). Pour L. F. Barrett (2009a) les états mentaux ne sont pas les éléments de base du cerveau. Ils sont produits par des éléments plus élémentaires. Il faut différencier la fonction d'un mécanisme et la manière dont il est produit (« *The contents of a psychological state reveal nothing about the processes that realize it, in much the same way that a loaf of bread does not reveal the ingredients that constitute it.* », p 330). Les interactions entre les différents éléments élémentaires poussent à ce que l'objet étudié soit considéré comme à un niveau supérieur (Van Orden & Kloos, 2003). Ainsi, pour L. F. Barrett, un état psychologique identique n'implique pas un état neuronal identique. Avec la théorie de « l'émergence » et la théorie de la « supervenience », la propriété mentale conceptualisée en psychologie n'est pas superflue, elle existe et est réelle mais on ne peut pas l'inférer à partir de l'état physiologique. Ces théories admettent qu'un état neuronal particulier donne un état mental particulier, même si ces neurones ne sont pas spécifiques avec l'état mental. Un état mental (état supérieur) dépend d'un état inférieur des neurones mais sans se

réduire simplement à celui-ci. Un état mental est donc supérieur à la somme de ces éléments constitutif, non-réductible aux états inférieurs et non-prédictible (non à cause de notre ignorance, mais parce que le contexte et les interactions produisent des états non-prédictibles). Le phénomène émergent a en conséquence des propriétés que les éléments inférieurs n'ont pas. La conscience par exemple ne peut être actuellement appréhendée que par cette théorie (Kievit et al., 2011b).

L. F. Barrett (2011) et bien d'autres ont ainsi recours à l'émergence et à la survenance pour expliquer les états mentaux. Le recours à l'émergence semble d'ailleurs important aujourd'hui pour expliquer des états mentaux complexes, plutôt que le recours à une vision modulaire et/ou localisable des états mentaux. Cependant, la survenance comme cadre explicatif n'est pas sans poser problème.

1.1.4.2 L'émergence robuste et la survenance

Le débat existe depuis 25 ans sur ces concepts d'émergence et de survenance. Les revues *Intellectica* (1997), *Syntheses* (2006), *Philosophical Studies* (2002) ont fait par exemple des numéros spéciaux sur ces concepts qu'il est impossible de rapporter in-extenso. Je renvoie à la thèse de Soubeyran (2012) pour une discussion actualisée de ce débat, travail très riche sur lequel je m'appuierai pour la suite de la discussion.

L'émergence se réfère à un processus par lequel de nouvelles structures et fonctions font leur apparition, dont les propriétés sont nouvelles, non réductibles aux propriétés des éléments microscopiques et non-prédictibles de façon réductionniste (à partir des éléments de ces éléments microscopiques). On résume souvent cela par « le tout est supérieur à l'ensemble des parties ». Dans le cas de la cognition, la notion d'émergence incarne l'idée que les propriétés mentales ne sont pas définissables dans les termes des propriétés physico-chimiques de leur support neurobiologique, et constituent ainsi un niveau de propriétés fonctionnellement autonome. La survenance est un cadre conceptuel permettant d'expliquer l'émergence de phénomène. Le concept de survenance (*supervenience*) est « *un concept introduit par Davidson pour exprimer une forme de dépendance sans réduction. Davidson considérait que les propriétés mentales dépendent (surviennent sur) des propriétés physiques, sans pour autant pouvoir être déduites de ces propriétés physiques. Pour lui, l'idée de survenance implique qu'il ne peut y avoir deux événements en tous points identiques physiquement mais différents mentalement, ou qu'un objet peut d'une façon ou d'une autre changer au niveau mental sans aussi changer au niveau physique* » (Van de Vijver, 1997,

p. 4). McLaughlin (2008, p. 13) rappelle la définition de la survenance où il y a une relation nomologique mais non-logique : « 1/ Si un ensemble de propriétés A résulte par survenance (*supervenies*) d'un autre ensemble plus fondamental de propriétés B, alors deux objets qui partagent des propriétés identiques de l'ensemble plus fondamental B ne peuvent différer dans les propriétés qu'ils partagent avec l'ensemble A. 2/ Si l'ensemble des propriétés A résulte par survenance de l'ensemble B, alors il se peut qu'aucune propriété de l'ensemble A ne puisse être définie ou reliée pratiquement (*manageably*) à aucun ensemble de propriétés de l'ensemble B. ». C'est une vision non-agrégative des éléments microscopiques pour caractériser les propriétés d'un élément macroscopique¹⁵.

1.1.4.3 Critique de la survenance

Le concept de survenance souffre de critiques. Jaegwon Kim est certainement le chercheur central dans la critique du physicalisme non-réductionniste faisant appel à l'émergence et à la survenance. Il me semble, qu'à la lecture de son ouvrage (J. Kim, 2009) et de différents articles, que Kim a développé une pensée critique en évolution et je ne prétends pas porter ici ses développements actualisés. Je me sers de certains de ses écrits afin de nourrir ma réflexion.

J. Kim (1999, 2006) se base sur une approche physicaliste sans adhérer au réductionnisme classique qui considère que tout est analysable à partir des éléments de bases de la matière (au niveau microscopique). Pour lui le point de départ dans une opération de réduction ne devrait pas consister à définir des corrélations entre propriétés de divers niveaux, c'est-à-dire à poser des lois-ponts (*bridge-laws*) entre les niveaux comme Nagel (1961) le propose. En effet, un grand nombre de système montrent que l'on ne peut réduire « logiquement » les propriétés d'un système macroscopiques aux propriétés microscopiques. Pour J. Kim, il faut pourtant expliquer la construction de nouvelles formes d'organisation à partir de parties plus simples. Il est nécessaire de rechercher les mécanismes qui puissent expliquer la nature et la nécessité de la connexion entre "propriétés" au niveau micro et "propriétés" au niveau macro. Pour J. Kim (2009, p. 41) : « *comment le mental peut-il être anomal (c'est-à-dire échapper au déterminisme) tout en étant causalement dépendant du domaine physique qui est gouverné par des lois déterministes strictes ?* ». J. Kim propose la "connexion conceptuelle" : toute fonctionnalisation présuppose une mise en relation conceptuelle entre la propriété

¹⁵ Je reprends cette terminologie « microscopique » et « macroscopique » usuelle dans la discussion sur l'émergence. Le niveau microscopique peut renvoyer aux neurones, et le niveau macroscopique peut renvoyer aux états mentaux dans notre discussion.

fonctionnelle émergente et les facteurs causaux qui la réalisent. J. Kim critique donc la survenance comme cadre donnant aux propriétés macroscopiques des fonctionnalités causales indépendantes des propriétés microscopiques. Pour J. Kim (1997), la théorie de la survenance n'apporte par exemple rien en physique ou en biologie, car tous les phénomènes présumés « survenir » trouve in fine des explications logiques avec le progrès de la science... ce en quoi il est critiqué (Van de Vijver (1997) pour des exemples). J. Kim discute de la clôture causale en physique (« *causal closure problem* ») qui renvoie à l'idée que tout effet physique a une cause physique suffisante, ce qui oblige les phénomènes d'un niveau supérieur à être en relations causales avec le niveau inférieur, il n'y a donc pas de place pour une causalité en dehors de ce niveau. Pour J. Kim (1997), si on considère que les phénomènes sont descriptibles à partir des éléments basiques (atomes), toutes les propriétés issues des systèmes supérieurs sont déterminées par ces éléments de base. Le point central de la discussion portée par J. Kim est la causalité descendante (*downward causation*) : avec la survenance, les propriétés émergentes influenceraient le système de base de façon indépendante du niveau micro. Pour Kim, si une propriété émergente M dépend d'un état basique P, et si M est la cause de modification P*, P peut être nomologiquement suffisant pour être qualifiée comme la cause de P*. Le phénomène émergent n'est rien de « supérieur » aux éléments de base et il n'y a pas de « survenance ». Si les propriétés émergentes influencent les propriétés de base et si les propriétés émergentes ne sont pas réductibles à des éléments de base, cela renvoie au dualisme cartésien : la « conscience » retrouve un rôle immatériel autonome car en dehors de la clôture physique.

La critique portée par Kim renvoie à de vraies questions sur la survenance comme base explicative telle que posée par L. F. Barrett (2011) et d'autres. Car en effet, l'émergence associée à la survenance semble donner une autonomie au niveau macro en contradiction avec le physicalisme. Ainsi, les exemples de la littérature (Cariani, 1997; Needham, 2009; Prosser, 2012 pour exemples) marquant que la réduction proposée par Kim n'est pas toujours fonctionnelle ne contredisent pas à mon avis la démarche critique de Kim envers la survenance. La critique essentielle envers l'émergence associée à la survenance peut se résumer en considérant que l'émergence n'est qu'un artifice explicatif (quelque chose de « magique », Kistler, 2007) pour donner un cadre explicatif à des phénomènes physiques trop complexes (actuellement) pour être intégrée à une théorie matérialiste partant des éléments de base. C'est une sorte de « boîte noire » (« *black box* ») sans aucun véritable ressort explicatif. Bien des exemples ayant fait appel dans le passé à la survenance comme facteur explicatif trouve aujourd'hui un cadre d'explication logique sans recours à ce concept. En effet, les

propriétés émergentes se ramènent souvent à « *est émergent ce qui ne peut être expliqué de manière réductive* » (Van de Vijver, 1997)). Cela revient à définir l'émergence d'une façon « négative ». Pour Soubeyran (2012), on parle d'émergence lorsqu'il y a : 1/ non-causalité partie-tout ; 2/ non-réductibilité du tout aux parties ; 3/ imprédictibilité ou non déductibilité du tout à partir des parties. L'émergence apparaît ainsi comme réponse à notre manque de connaissance (concept épistémique). « *L'émergentisme classique échoue à définir positivement la nature des relations entre les parties et le tout, entre les différents niveaux métréologiquement hiérarchisés. Cet échec semble réduire l'émergence à un concept purement épistémique* » (p. 65).

Cependant, pour bien des auteurs, sur le terrain de la conceptualisation scientifique, la caractérisation d'émergence est souvent nécessaire. Sinon, cela revient à considérer que la seule cause de tout sont les particules élémentaires en physiques (Bontly, 2002), ce qui n'est pas très constructif pour un champ scientifique (par exemple cela reviendrait à considérer pour Bontly que le tabac ne provoque pas le cancer car la cause du cancer reviendrait aux particules élémentaires) ; et même pour Humphreys (1997b) cela laisse considérer que toutes les théories (chimique, biologique, ...) sont erronées car ne font pas appel aux particules élémentaires. De plus, si des propriétés émergentes (comme la température) sont réduites aux éléments de base (cinétique de molécules), on perd ces notions de généralisation (la propriété causale de l'élément supérieur n'équivaut pas à la propriété causale de l'élément inférieur, car une même température fait correspondre pour différents gaz à différentes pressions des propriétés cinétiques différentes). La réduction pose donc problème car le niveau supérieur est nécessaire pour conceptualiser les phénomènes observés sans devoir avoir recours aux particules élémentaires pour construire un modèle explicatif (Bontly, 2002). Mais il me semble que cette critique de la réduction ne remet pas en cause la critique de la survenance telle que posée par J. Kim. Kistler (2007) : « *Une autre manière de rendre une réduction plus acceptable est de faire remarquer que la réduction d'un concept ne conduit pas à son élimination, et à plus forte raison ne conduit pas à contester la réalité du phénomène qui faisait l'objet du concept réduit.* ». Pour J. Kim, si une propriété macroscopique donnée survient toujours sur un même ensemble de propriétés microscopiques, on peut postuler que cette propriété macroscopique est déterminée par ces micro-propriétés. Pourtant selon J. Kim, il faut distinguer détermination et explication. Expliquer réductivement une certaine propriété macroscopique supposerait de pouvoir l'expliquer uniquement sur la base des propriétés de ses parties, c'est-à-dire de pouvoir la dériver logiquement ou conceptuellement à partir de ces propriétés. La fonctionnalisation de la propriété à réduire doit permettre de l'identifier à la

propriété de premier ordre qui la réalise. La propriété du phénomène émergent devient une propriété de second ordre des éléments microscopiques, c'est une propriété « macroscopique microbasée » : « *Une propriété microbasée est constituée de microconstituants - c'est-à-dire constituée par les microparties de l'objet qui a cette propriété, et par les propriétés et relations caractérisant ces parties. Mais nous devons dire clairement que de telles propriétés microbasées sont des macropropriétés, et non des micropropriétés.* » (cité par Soubeyran, 2012, p. 229). Pour J. Kim, les systèmes considérés comme émergents ne sont pas qu'un simple agrégat de ces éléments de base, ils ont des propriétés non-prédictibles au niveau élémentaire, ils ne sont pas réductibles aux éléments de base car s'il y a une prédictibilité inductive (je constate qu'un élément de base P provoque une propriété émergente M), une prédictivité déductive n'est pas possible (connaissant toutes les propriétés de P je ne peux en déduire les propriétés de M). Cependant pour J. Kim on doit pouvoir considérer le « réalisateur » microbasé de la propriété macrobasée. Il faut donc avoir une approche réductionniste sans articuler forcément des « lois pont » mais à partir de la « réduction fonctionnelle » adaptée aux phénomènes émergents.

1.1.4.4 Les systèmes non computationnels

Est-ce que le réductionnisme fonctionnel de J. Kim implique de devoir se passer du concept d'émergence pour appréhender les propriétés mentales ? En fait, la discussion sur l'émergence a essentiellement portée jusqu'à récemment sur un concept d'émergence considéré comme *robuste*. L'émergence « robuste » postule un certain nombre de propriétés à un phénomène émergent telles que : son irréductibilité, son imprédictibilité, sa complexité organisationnelle, sa nouveauté, ou encore son pouvoir causal propre. Cette approche est contradictoire avec la clôture causale du domaine physique qui implique que tout phénomène physique puisse faire l'objet d'une micro-explication. Sur le plan méthodologique, l'explication scientifique semble devoir donc être nécessairement réductionniste. Le cœur de la discussion critique de J. Kim porte sur la survenance, c'est à dire la causalité propre des propriétés émergentes indépendantes du niveau micro. Raison pour laquelle Humphreys critique la position de J. Kim en considérant qu'il y a de fait des niveaux qualitativement différents et donc émergence ; mais Humphreys appelait à utiliser l'émergence sans recourir à la survenance (Humphreys, 1997a, 1997b) ; c'est-à-dire à considérer des niveaux macro comme non-réductible aux niveaux micro tout en ne considérant pas pour autant que les niveaux macro aient une autonomie causale indépendante des niveaux micro.

De fait, une famille de modèle semble caractériser une émergence sans survenance : des systèmes dynamiques non-linéaires (souvent connus par la « théorie du chaos »). Depuis les travaux de Poincaré on sait que des systèmes dynamiques peuvent être sensibles aux conditions initiales, et il est impossible pour un grand nombre de systèmes dynamiques de prédire leur trajectoire de façon « logique », c'est-à-dire par un formalisme mathématique (il n'y a pas de règle qui relie le niveau micro et le niveau macro). On ne peut donc prédire leurs trajectoires par les simples propriétés des éléments microscopiques car la seule solution pour prédire la trajectoire de tels systèmes est de simuler l'ensemble des trajectoires possibles, on parle de « l'irréductibilité computationnelle ». Un système physique est computationnellement irréductible « *si son comportement ne peut être déduit d'une procédure prédictive plus courte que la simulation ou l'observation de l'évolution du système* », définition tirée de Soubeyran (2012, p. 216). Le besoin de recourir à une simulation n'est pas dû à l'état de notre connaissance car il n'existe pas de description plus courte de l'évolution de tels systèmes que la description de chacun des états successifs du système au cours de cette évolution. L'explication est logiquement incompressible (l'émergence n'est donc pas un concept épistémique). De plus, la structure et les propriétés des états stationnaires de ces systèmes, non-prédictibles logiquement, ne semble pas pouvoir être conceptuellement connectées aux propriétés microphysiques de ces systèmes (ce ne sont pas des propriétés macroscopiques conceptualisables par des micro-propriétés). Comme la réduction logique et conceptuelle au niveau microscopique est impossible, il est impossible de recourir à la réduction fonctionnelle proposée par Kim pour de tels systèmes (Kistler, 2007). La fonctionnalisation d'une propriété macroscopique telle que proposée par Kim doit en effet permettre de l'identifier à un mécanisme microscopique, ce qui est impossible avec bien des systèmes dynamiques non-linéaires.

1.1.4.5 L'émergence faible

L'émergence robuste comme mode explicatif de systèmes non-dynamiques renvoie à la survenance... et buterait sur les problèmes posés par J. Kim : l'argument de l'exclusion causale (selon lequel il y a tout au plus une seule explication indépendante complète et entière pour chaque événement). Mais des systèmes dynamiques, comme certains systèmes non-linéaires, ont un mode explicatif qui fait appel à une forme d'émergence. Cette émergence se différencie de l'émergence robuste car le niveau explicatif reste microscopique, et ne fait plus appel à une causalité macroscopique différente des états microscopiques (on reste dans le

cadre de la clôture physique) ; cependant la structure macroscopique n'est réductible au niveau microscopique ni logiquement ni conceptuellement. Kistler (2007) reprend l'idée qu'il n'y a vraisemblablement pas de propriété épistémiquement émergente (au sens épistémique, une propriété est émergente si nous sommes incapables de l'expliquer) car toutes les autres propriétés macroscopiques ne sont épistémiquement émergentes que provisoirement, c'est-à-dire relativement à un certain cadre de connaissances et de théories ; par contre il y a des propriétés ontologiquement émergentes, c'est-à-dire des « *propriétés qualitativement différentes de toutes les propriétés que peuvent posséder leurs parties* » (p. 26). Ce concept d'émergence est compatible avec la réduction : « *la propriété P d'un système complexe est ontologiquement émergente si 1) elle est systémique, c'est-à-dire aucune partie du système ne la possède, 2) elle est qualitativement différente des propriétés que possèdent les parties du système, 3) elle respecte le physicalisme, c'est-à-dire elle est intégralement déterminée par les propriétés des parties du systèmes, ainsi que par leur interactions mutuelles et avec l'environnement, et 4) elle est stable ou « robuste », c'est-à-dire invariante par rapport à des petites perturbations dans les propriétés microscopiques sous-jacentes.* » (p. 26-27). Ce concept d'émergence ontologique renvoie au concept d'« émergence faible » (« *weak emergence* », Bedau (1997, 2008)). Un état émergent au sens faible est un état macroscopique qui peut être dérivé de la connaissance de la micro dynamique du système, mais seulement via la simulation ou la modélisation de l'ensemble des interactions micro-causales qui ont conduit à ce macro-état, Bedau (2008): « *Si P est une macro-propriété d'un certain système S, alors P est faiblement émergent si et seulement si P est explicable à partir de tous les micro-faits préalables de S mais seulement d'une manière incompressible.* ».

Cependant la taxinomie de l'émergence est en construction et je n'en donne qu'une vision simplifiée, je renvoie à Humphreys (2008) pour avoir une photographie des différentes formes d'émergences posées par la littérature philosophique. Il me suffit pour la suite de mon développement de poser les systèmes dynamiques complexes comme des systèmes amenant une forme d'émergence, c'est-à-dire où les propriétés macro sont non-réductibles conceptuellement et logiquement aux propriétés micro ; mais émergence qui ne fait pas appel à la survenance.

La non-réduction logique au niveau microscopique est, avec l'émergence faible, non-contradictoire avec une causalité des éléments microscopiques. Cette approche relie les micro-explications au phénomène émergent et contourne le problème de la clôture causale en physique posé par J. Kim. De plus, la structure macroscopique d'un phénomène peut être conceptualisée dans un langage simple ; alors que le recours au niveau microscopique

complexifie cette conceptualisation (car le seul mode explicatif conceptuellement est de décrire toutes les trajectoires possibles à partir des états microscopiques). Nous retrouvons donc avec l'émergence faible un cadre intégrant la critique sur le terrain de la conceptualisation scientifique de la théorie réductionniste (Bontly, 2002), car la conceptualisation du niveau macro n'est pas réductible au niveau micro. Les états macroscopiques sont donc qualitativement différents des états microscopiques, le tout étant « supérieur à l'ensemble des parties », d'où la caractérisation émergente des telles structure macroscopiques. Mais à la différence de l'émergence robuste, ces systèmes dynamiques n'ont pas recours à une « boîte noire » explicative contradictoire avec l'hypothèse de la clôture physique, telle que l'émergence robuste le propose.

1.1.5 La propriété mentale comme qualité émergente ?

La propriété mentale renvoie à un phénomène particulier car elle rend compte de processus complexes non-observables débouchant sur des attracteurs non prédictibles. La réalité tangible des propriétés mentales, comme formes émergentes de tels processus complexes, renvoie donc au problème de leur formalisation qui ne peut partir des processus en amont car une propriété émergente ne peut être réduite aux processus la générant.

1.1.5.1 La propriété mentale comme phénomène émergent

Wallot & Van Orden (2011) considèrent qu'un système « *compenent-dominant* » a des propriétés émergentes faibles et le résultat peut être déduit des propriétés causales des composantes et de leur arrangement ; à l'inverse un système « *interaction-dominant* » a des propriétés émergentes fortes et le résultat ne peut être alors déduit des propriétés causales des composantes. Les résultats semblent montrer que le comportement humain a une dynamique « *interaction-dominant* » (Van Orden et al., 2011). Ainsi, le système ne peut être réduit à ses parties, il y a émergence (Geert, 1994). Les neurones pourraient être assimilés à des systèmes dynamiques non-linéaires, computationnellement irréductibles, et avec des états macroscopiques (les propriétés mentales) caractérisables comme phénomènes émergents (au sens faible). Barandiaran & Moreno (2008) : « *L'auto-organisation des impulsions neuronales à différentes échelles résulte dans la génération de structures intermédiaires, lesquelles interagissent elles-mêmes et aident à organiser les dynamiques globales du système dans un sens plus large que les seules interactions locales des impulsions. Ceci nous oblige à distinguer les niveaux dynamiques micro et macro (et diverses hiérarchies de niveaux) et*

nous permet de parler de fonctionnalité émergente (processus neuronaux microscopiques dont la contribution à l'auto-maintenance du système est marquée par l'émergence de structures qu'ils constituent ou auxquelles ils participent). ». Huetz (2006) » « L'activité neuronale est maintenant considérée comme un système dynamique dont les évolutions sont décrites par des systèmes d'équations différentielles [...] Les composants du système (les neurones) sont interconnectés de sorte que l'ensemble puisse réaliser, à l'échelle globale, différentes configurations d'activation identifiées à des 'états' cognitifs. » (p. 12).

L'approche par les systèmes complexes non linéaires permet, via l'émergence faible, de contourner le dilemme entre une exigence non-réductionniste pour comprendre les états psychologiques comme d'une nature qualitativement différents des états neuronaux, tout en conservant une orientation matérialiste/physicaliste de ces états mentaux. Avec les *Attractor Neural Networks*, le réseau de neurones est envisagé comme un système dynamique non-linéaire complexe car les influx nerveux ne se font pas de façon linéaire et discrète. La dynamique au niveau macro du réseau de neurones ne semble pas computationnellement réductible. De plus, la fonction de la structure macroscopique (les attracteurs du système dynamique assimilables à des propriétés mentales) ne peut être réduite conceptuellement aux propriétés du niveau micro (les neurones et processus cognitifs de base). Les propriétés mentales peuvent donc être caractérisées comme des systèmes complexes dynamiques « interaction-dominant » (en interaction permanente avec des stimuli internes et externes au système) et donc comme non-prédictibles (car computationnellement irréductibles) et ayant un niveau conceptuel (les propriétés mentales) différent des réseaux de neurones ou processus cognitifs en amont. Nous conservons ainsi une approche physicaliste (les propriétés mentales découlent des états du réseau de neurones) sans pouvoir réduire l'un à l'autre ni au niveau conceptuel ni au niveau logique. Le réseau de neurones ainsi modélisé correspond à la caractérisation d'un système complexe avec des phénomènes émergents (au sens émergence faible tel que défini précédemment). Les états mentaux, comme états stationnaires (attracteurs) d'un réseau complexe non-linéaire, conservent leurs réalités physiques mais ne sont conceptualisables qu'à un niveau supérieur et non-prédictible à partir d'une modélisation microscopique.

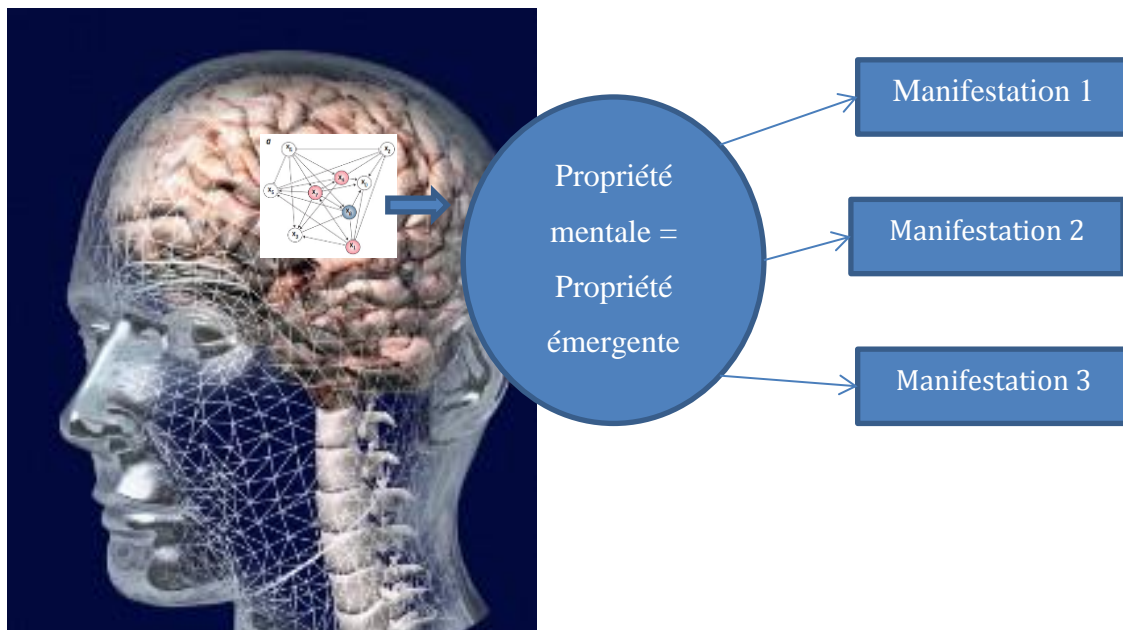
Le recours à l'émergence et à la survenance me paraît problématique tel qu'abordé par L. F. Barrett (2009a, 2011) ou d'autres car l'émergence introduit une boîte noire explicative. L'émergence faible telle que caractérisée ci-dessus semble permettre une démarche « réaliste », sans réduire les propriétés mentales au niveau microscopique, mais sans poser l'émergence comme un simple artifice explicatif.

La qualité « émergente » qui est caractérisée par la propriété mentale permet de déplacer le problème du rapport à la « réalité » des propriétés mentales. Pour amorcer notre discussion, je reprends les mots de Soubeyran (2012, pp. 236-237) « *Nous avons montré que cette organisation macroscopique ne pouvait être logiquement déduite à partir des propriétés des composants microscopiques, prédite sans décrire ou simuler l'ensemble des micro-étapes qui ont conduit à la macro-organisation. Pourtant, la cohésion du système, son organisation spatio-temporelle robuste, permet de décrire sa structuration macroscopique de façon simple. L'émergence correspond ainsi à une régression inattendue de la complexité entre la description des interactions microscopiques et la description macroscopique. Le vocabulaire de cette description macroscopique apparaît comme compositionnellement irréductible à celui de la description microscopique, les prédicats des deux descriptions ne semblent pas pouvoir être corrélés. Les propriétés dynamiques globales qui caractérisent les systèmes émergents ne semblent donc pas pouvoir être conceptuellement connectées aux propriétés microphysiques de ces systèmes.* »

Les neurones en tant que systèmes dynamiques complexes peuvent générer des propriétés non-réductibles au réseau de neurones. Ces propriétés doivent se conceptualiser dans un vocabulaire différent de celui des micro-propriétés. Cette qualité émergente n'est donc pas conceptualisable à partir des processus cognitifs multiples qui interagissent. Il n'y a pas de correspondance entre le concept biologique et le concept psychologique (Miller, 2010). Le « rouge » par exemple est un concept complètement différent entre l'activité neuronale et la perception expérientielle de cette couleur. La conceptualisation de l'objet mesuré en psychologie (propriété mentale) ne peut donc renvoyer qu'à ce que la psychologie conceptualise sans recours possible à la neuroscience.

Si nous comprenons les construits comme des propriétés émergentes, nous pouvons les schématiser comme suit :

FIGURE 9 : FIGURE DE LA PROPRIÉTÉ ÉMERGENTE D'UN NEURAL NETWORKS



1.1.5.2 Au-delà des neurones

Cependant, cette schématisation me paraît restreindre le processus réel. De prime abord, en accord avec Varela et al. (1999), la discussion académique isole le plus souvent le cerveau sans inscrire corporellement l'activité mentale ; or il me paraît nécessaire d'avoir une vision corporelle globale de l'activité mentale. L'intérêt qui semble émerger pour l'intestin par exemple témoigne de cette nécessité (Enders, Enders, & Liber, 2015). L'inscription corporelle de l'activité mentale renvoie la personne à son existence en tant que sujet.

Cette inscription corporelle renvoie à « l'être » au-delà de sa simple entité neuronale. Il semble que bien des travaux se focalisent sur la personne en tant qu'être biologique sans intégrer l'être social au sens des interactions sociales propres à chaque individu à chaque instant. Je rappelle que les processus sous-jacents à la propriété mentale émergente est environnement-dépendant, interaction avec le contexte et l'environnement (Van Orden et al., 2011). Kendler (2012) considère ainsi qu'il faut, pour appréhender une propriété mentale, intégrer différentes catégories qui vont au-delà de la biologie et renvoie à des aspects sociaux : génétique, neurochimie, structures neuronales, biologie, personnalité, traumatismes personnels, interactions sociales, politique, culturels. Pour Huetz (2006) « *L'émergence de telle ou telle configuration dépend des relations locales entre les éléments du système et celles-ci dépendent des procédures d'apprentissage et des modifications ultérieures.* ».

Mais au-delà de l'apprentissage, il faut considérer la propriété émergente comme résultante d'un processus interne et externe permanent. Varela et al. (1999) critiquent la vision de « l'esprit humain » déconnectée de toute référence au sujet. Varela (1998) « [On ne peut] *pas comprendre la cognition si on l'abstrait de l'organisme inséré dans une situation particulière avec une configuration particulière, c'est-à-dire dans des conditions écologiquement situées* ». Varela et al. (1999) en appellent aussi aux attracteurs et à l'émergence (page 136 de leur livre) comme des états stabilisés de cette interaction. Je reprends la synthèse proposée par Huetz (2006) « *Dans le cadre d'une étude de la cognition, le sujet ne peut pas être considéré comme un système passif de perception de stimuli externes [...] La cognition ne repose donc pas sur des représentations symboliques du monde extérieur mais se rapporte plutôt à un processus autonome résultant de l'interaction du sujet avec son environnement [...] Les processus cognitifs sont le résultat de ce couplage perception-action. L'activité cognitive d'un sujet est dans ce sens indissociable de l'expérience du sujet [...] l'étude de l'activité neuronale ne peut se passer de la prise en compte de l'expérience du sujet* » (p. 13-14). Ou pour reprendre Juhel s'appuyant sur Lamiell pour décrire la complexité de l'unité du sujet relatif à l'action (cité par Juhel, 2014) « *L'individu est un unitas multiplex, "une entité qui, bien que composée de multiples parties, forme une unité singulière et constitue en tant que telle, en plus de ses unités en fonctionnement, un être unitaire, auto-actif et dirigé par un but"* ».

Je m'inscris dans cette vision intersubjective (ou interrelationnelle¹⁶) des personnes, appréhendant les propriétés mentales comme interdépendantes de l'environnement de la personne. Un individu n'existe que dans un espace social, par une pratique permanente de son environnement et dépendant en permanence de son environnement. On ne peut réduire la personne à l'être biologique isolé. Sa nature psychologique ne peut être appréhendée que relativement à cette interaction permanente avec le cadre social de sa vie. Comme le rappelle Nicodème (2015), se référant à l'introduction à la critique de l'économie politique de K. Marx, « *l'individu est inscrit socialement, il n'est pas ontologiquement un* ». On retrouve donc une critique ancienne de Marx envers le matérialisme tel que développé par certains philosophes au XIX^{ème} siècle. L'homme est avant tout « *un produit social* », et Marx (1859/1972) le caractérise de la sorte « *l'homme est dans le sens le plus littéral un animal politique, non seulement un animal sociable, mais un animal qui ne peut s'isoler que dans la*

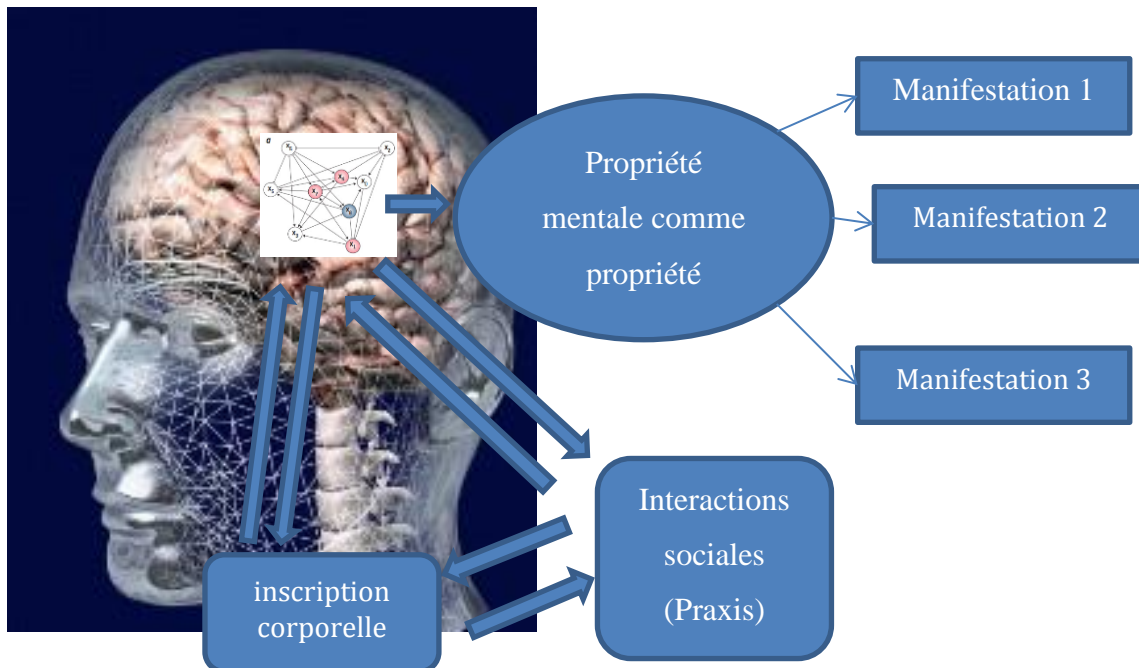
¹⁶ P. Corcuff m'a gentiment expliqué qu'en sociologie on utilisait le concept d'interrelation aujourd'hui à la place de l'intersubjectivité. Je considère que pour mon travail cette différence conceptuelle n'est pas très importante.

société » (p. 150). Ainsi, pour Marx « le principal défaut de tout matérialisme jusqu'ici [...] est que l'objet extérieur, la réalité, le sensible ne sont saisis que sous la forme d'Objet ou d'intuition, mais non en tant qu'activité humaine sensible, en tant que praxis¹⁷, de façon subjective [...] Feuerbach veut des objets sensibles, réellement distincts des objets de la pensée : mais il ne saisit pas l'activité humaine elle-même en tant qu'activité objective » (Première thèse sur Feuerbach, Marx & Engels, 1846/1972).

Il faut donc appréhender la propriété mentale comme propriété émergente d'une praxis humaine générant des interactions continues avec son environnement, et non considérer la personne comme un isolat à appréhender comme objet biologique.

L'émergence d'une propriété mentale devrait donc plutôt se représenter par ce schéma :

FIGURE 10 : FIGURE DE LA PROPRIETE EMERGENTE D'UN NEURAL NETWORKS INSCRIT DANS UNE PRAXIS



1.1.5.3 Propriété émergente réelle

Le recours à l'émergence est usuel avec les variables latentes. Mais dans bien des articles cette émergence ne semble pas correspondre à une réalité, mais plutôt à une heuristique

¹⁷ Le terme « praxis » est remplacé dans l'édition par « pratique », mais dans le livre de Engels « Ludwig Feuerbach » (Engels, 1888), le terme « praxis » est mis. Il me semble plus pertinent de conserver « praxis » car il implique une dimension moins réductrice qu'une simple pratique.

méthodologique pour résumer un système complexe. Pour illustrer cela, je prends pour exemple Willoughby (2014) qui discute des fonctions exécutives (EF), sa discussion peut s'étendre pour lui à (*inhibitory control, working memory, attention shifting*). Pour lui le modèle avec variable latente ne capte pas la réalité émergente, et la modélisation avec variable latente n'est qu'une « approximation » de cette complexité : « *To the extent that EF is an emergent property that is defined by interactions between neural systems, [...] To be clear, although neither formative nor reflective latent variables may adequately capture the variety or complexity of cognitive processes that support goal-directed behaviors, the suggestion here is that formative latent variables may provide a better approximation to modern theoretical accounts of EF than do reflective latent variables. Similar to our consideration of Jarvis et al.'s (2003) plausibility questions in the focus article, appealing to differences in the presumed ontological status of latent variables is a necessarily subjective process for which no definitive conclusions can be drawn » (p. 177). Dans cet article, la propriété mentale n'émerge pas comme une « réalité », mais comme une approximation. Ces auteurs caractérisent les propriétés de bases comme des modules, et de ce point de vue la propriété émergente n'est qu'une heuristique nécessaire. Cet article symptomatise à mes yeux la contradiction usuelle dans la littérature académique entre les processus de bases et la propriété émergente. Le passage des modules de base à l'émergence (comme Willoughby, Holochwost, Blanton, & Blair, 2014) pose un problème dès que l'on part des modules de base car alors la propriété émergente est un artifice explicatif.*

Pourtant toute la littérature pose les propriétés mentales comme réelles. L'approche en *Neural Networks* avec des états stabilisés rompt avec la vision d'un simple processus à appréhender en tant que processus. C'est l'état stabilisé, produit par des interactions multiples internes et externes à la personne, qui génère une émergence de tous ces éléments en une propriété mentale. Borsboom, Kievit, Cervone, & Hood (2009) le résument très bien dans cet exemple (même si le recours à la *supervenience* est critiquable) où l'instanciation d'une propriété est multiréalisable neuro-biologiquement : « *Nonreductive physicalism roughly holds that psychological states (like, for instance, 'believing that π is not a rational number')* can be realized in an infinite number of ways in the human brain. Thus, although there cannot be differences in psychological states if there are no differences in the physical structures that realize them (supervenience and materialism), there may be differences in the physical structures that serve to realize the same psychological state (multiple realizability). » (p. 24). Il y a bien une réalité de ce π dans le cerveau, réalité d'une propriété mentale (et sociale en

même temps car π est aussi un objet social, je vais y revenir) émergent au sens où cette propriété n'est pas formalisable en terme neurobiologique.

La régression du complexe est réelle par une propriété émergente (attracteur du réseau de neurones). On ne peut revenir au niveau micro (les neurones ou les processus cognitifs en amont). Il faut choisir le niveau d'analyse : analyser les éléments micro générants cet émergence (les processus de base), mais alors il faut aller au-delà de ces processus cognitifs de bases et appréhender la personne dans sa complexité sociale, et donc proposer un modèle de relations complexes ; ou analyser cette propriété émergente comme une réalité émergente, et donc partir de ce niveau sans prétendre redescendre au niveau en-dessous.

1.1.5.4 Emergence et variables latentes

Le processus de réduction du complexe lié à l'émergence oblige de poser la conceptualisation de la propriété à un autre niveau que les processus le générant. La réalité de la propriété mentale ne peut pas se matérialiser physiquement au sens neuro-biologique du terme mais correspond à une « *organisation spatio-temporelle robuste* » (Soubeyran, 2012). Sa réalité, en reprenant J. Kim (2009) pour qui « *être réel, c'est avoir des pouvoirs causaux* », est caractérisée par des manifestations. Je ne développerai pas la causalité qui renvoie à un champ de la philosophie des sciences très discuté¹⁸, mais nous pouvons considérer « réelle » une propriété émergente causant des « manifestations ». Pour Borsboom et Cramer par exemple la propriété mentale est appréhendée par ses manifestations : « *the proposition that mental disorders cannot be separated from their symptoms* ». Comme la propriété mentale est non observable comme objet neurobiologique, et que les propriétés mentales génèrent des manifestations, c'est au niveau de ces manifestations qu'il faut essayer de l'appréhender. Une telle approche des propriétés mentales donne une légitimité scientifique aux modèles avec variables latentes pour inférer des propriétés mentales à partir de leurs manifestations, sans pour autant y voir un cadre instrumentaliste.

En conséquence, la démarche avec les variables latentes ne doit pas essayer de lier ces propriétés mentales, d'un niveau macro, à des propriétés d'un niveau micro (les neurones ou processus cognitifs de bases) comme l'ont tenté différents travaux, car la propriété émergente est irréductible aux processus micro qui l'ont créée. Il faut partir du niveau de manifestation de la propriété émergente (le niveau macro). L'approche des propriétés mentales comme des

¹⁸ Pour la « causalité » et modèles avec variables latentes, je renvoie au chapitre de Juhel qui permet d'appréhender ce problème de façon actualisé : (Juhel, 2015)

propriétés émergentes de processus neuronaux complexes permet de contourner le problème de la « boîte noire » causale des mesures effectuées qui semblait rendre caduc l'usage potentiel des variables latentes (comme le questionnaient Van der Maas et al., 2011) « *the lack of a psychological justification for this key property of the measurement model compromises test validity. The reason is that validity requires a causal mechanism linking the trait or ability with the item responses (Borsboom & Mellenbergh, 2007; Borsboom et al., 2004). In the absence of such a mechanism, the relation between the targeted attribute and the item responses is essentially a black box, and the psychological appropriateness of the function that describes this relation becomes an article of faith.* » (p. 353). La propriété mentale n'est pas « un processus cognitif » (complexe) mais un objet en soi, considéré comme phénomène émergent de ce processus. On ne peut pas chercher à le lier de façon épistémique aux processus neuronaux l'ayant généré.

C'est parce que la réalité de la propriété émergente n'est pas clairement définie par bien des auteurs que ces derniers se fourvoient à mon avis, soit en essayant de lier le niveau micro (les processus élémentaires) et le niveau macro (la propriété émergente) ; soit en se détournant des variables latentes en conséquence de cette impossibilité de lier les processus complexes (en amont) aux propriétés mentales. Par exemple des modèles d'intégration de la complexité, sans utiliser la régression de ce complexe par des variables latentes, sont discutées dans l'ouvrage de Markus & Borsboom (2013a), notamment le modèle avec liens, « *bond modele* » (Bartholomew et al., 2009). Ces auteurs basent leur travail sur celui de Thomson (1916) et ils discutent qu'il y a une confluence de causalité. Ces auteurs ne posent pas le problème de la gestion de la complexité même s'ils évoquent l'interaction entre différents réseaux comme Thomson l'avait suggéré. Or, fondamentalement, ce que Thomson écrit (1951) ne semble pas contradictoire avec une vision de la complexité vs régression de la complexité nécessaire : « *What the "bonds" of the mind are, we do not know. But they are fairly certainly associated with the neurones or nerve cells of our brains, of which there are probably round about ten thousand million in each normal brain. Thinking is accompanied by the excitation of these neurones in patterns. The simplest patterns are instinctive, more complex ones acquired. Intelligence is possibly associated with the number and complexity of the patterns that the brain can (or could) make.* » (cité par Bartholomew et al., 2009 p. 569).

Ainsi Markus et Borsboom, en ne discutant pas de la nature réelle des propriétés mentales, laissent entendre que les approches basées sur les réseaux par exemple pourraient être plus efficaces que les variables latentes. Or ces approches essaient d'intégrer la complexité des processus cognitifs dans le cadre des variables latentes, et posent en fait des approches

discutant de deux niveaux différents de la réalité : les processus cognitifs comme réseau dynamique complexe, et le phénomène émergent de ce réseau. Si on veut analyser le phénomène émergent d'un système complexe, il faut poser le modèle conceptuellement et méthodologiquement à ce niveau émergent, sans recours au réseau l'ayant généré. Cela n'implique bien évidemment pas que les travaux sur les processus cognitifs par neuro-imagerie par exemple soient contradictoires avec les travaux avec les variables latentes, mais en tant que miroir les uns des autres mais non-conjointes car discutant à des niveaux différents de la réalité.

1.1.6 Conclusion sur l'ontologie des propriétés mentales

L'approche des propriétés mentales comme des propriétés émergentes (au sens de l'émergence faible) de processus neuronaux complexes, formalisés par des variables latentes, permet de contourner bien des problèmes posés par la littérature. Je repars des *Network Analysis* pour comprendre les apports de la modélisation par variables latentes.

1.1.6.1 Conclusion sur les *Network Analysis*

Le problème central de l'approche *Network Analysis*, comme déjà dit, est l'incapacité de définir clairement la propriété mentale et de n'user que des manifestations en retirant toute « existence » à la propriété mentale. A vouloir modéliser le complexe, on supprime du modèle l'entité que l'on veut modéliser : la propriété mentale. L'approche *Network Analysis* à l'intérêt d'intégrer en quelque sorte l'interdépendance du sujet avec son environnement par l'intégration de manifestations interreliées, et en ceci c'est une réelle avancée dans la littérature académique qui considère trop souvent l'individu comme un objet biologique isolé. Mais cette approche n'intègre pas l'objet étudié lui-même (la propriété mentale). On nous propose une nouvelle complexification de la réalité sans réduction de celle-ci car la propriété mentale étudiée disparaît en tant qu'objet d'étude. Leur modélisation du complexe va conduire demain à devoir considérer la propriété mentale comme une forme émergente du réseau de ses manifestations pour gérer cette complexité ; on n'aura plus une propriété émergente d'un réseau (cognitif) en amont mais une propriété émergente d'un réseau (de manifestations) en aval de la propriété mentale.

Je suis en parfait accord avec Borsboom & Cramer (2013) qui considèrent, comme je l'ai déjà écrit, qu'en psychologie nous avons rarement une relation claire partant du biologique (un organe malade analysable en tant que tel) pour arriver aux symptômes. Comme je l'ai dit plus

haut, pour Borsboom et Craemer la propriété mentale est appréhendée par ses manifestations (« *the proposition that mental disorders cannot be separated from their symptoms* »). Mais considérer cela n'interdit pas de poser l'existence d'une véritable propriété à analyser en tant que telle. Visiblement Borsboom et d'autres essayent via les *Network Analysis* d'appréhender la complexité liée aux propriétés mentales. Mais l'approche *Network Analysis* confond à mes yeux, dans sa genèse, la complexité sous-jacente des processus générant une propriété mentale émergente et la propriété mentale émergente qui ne forme pas un réseau complexe mais est à appréhender comme une propriété en soi. Même en posant les manifestations comme réactives avec la propriété mentale comme dans les *Network Analysis* (« *reactive indicators* », Markus & Borsboom, 2013a), ce qui est certainement valide, nous pouvons considérer devoir appréhender une propriété émergente comme correspondante à un état stabilisé d'un système « robuste » (Soubeyran, 2012) qui intègre dans sa stabilité les interactions internes et externes à la personne. La notion de robustesse de la propriété émergente considère que des variations (faibles) des processus (interne ou externe à la personne) ne modifient pas l'attracteur.

Poser la propriété mentale comme une propriété émergente, donc caractérisable par ses manifestations ne la réduit pas à un réseau de ses manifestations comme le fait le *Network Analysis*. La propriété mentale est une entité réelle de la personne, comme l'affirment par ailleurs les chercheurs développant les *Network Analysis*. Elle est à considérer en tant que telle, certes réifiée intersubjectivement (je reviendrai sur ce point par la suite), mais existant réellement. Donc on doit essayer de l'appréhender comme un objet, c'est-à-dire la conceptualiser et la mesurer (je reviendrai sur la mesure dans une section suivante) à partir de ses manifestations, seules éléments tangibles permettant de la caractériser.

Cependant, au-delà de l'intérêt de voir la personne de façon dynamique, l'apport des travaux sur les *Network Analysis* me paraît réel pour mieux caractériser des propriétés mentales. Les quelques travaux que j'ai lus sur les *Network Analysis* se basent sur des propriétés mentales discutées et critiquées dans la littérature comme des traits (la personnalité ou la dépression par exemple). La critique des travaux ayant formalisé en un trait la personnalité ou la dépression a donné lieu à des publications importantes, et peut être que les *Network Analysis* permettent de confirmer une réalité plus complexe que cette formalisation en simple variable latente unidimensionnelle de la personnalité ou de la dépression (ou en 5 traits pour la personnalité, voire par exemple Juhel (2014) qui considère « *Les systèmes de personnalité et les contextes sociaux ont des effets interactifs et mutuellement dépendants sur le comportement.* »). Pour conserver une approche réaliste, où une propriété mentale existe et se manifeste, il faut

pouvoir considérer que nous avons un phénomène étudié latent unique, je reviendrai par la suite sur ce point.

De ce fait je suis en accord avec Cramer et al. (2012, p. 429) sur la critique de l'évidence de l'usage des variables latentes telle que pratiquée aujourd'hui, comme pour la personnalité et la dépression par exemple (« *That is, we do not object to latent variable 'techniques' but we do object to a latent variable 'theory' in which the measurement model with a common cause structure is interpreted as evidence for latent causal entities that operate in the minds of individuals, causing all sorts of cognitive, affective and behavioural patterns (see also Borsboom, 2008b).* ».) Mais comme le rappelle Borsboom (2008b), la nature latente d'une variable n'est pas intrinsèque à la propriété étudiée, mais à son accessibilité épistémique (il y a la confusion dans ce que dit Borsboom par la suite entre « variable » et « propriété » car une variable n'est pas « observable » ou « non observable », c'est la propriété que représente la variable qui est observable ou non, mais indépendamment de ce point son développement me paraît pertinent) : « *a latent variable model is simply a model that relates the expectation of observables to a latent structure through some regression function [...] Observed variables are variables that are somehow epistemically accessible to the researcher, whereas latent variables are not epistemically accessible.* » (p. 26). Pour lui le sexe ou l'âge ne sont pas plus observable qu'une propriété mentale. Une propriété mentale n'a pas de « *surplus meaning* » : « *First, it is misleading to take the distinction between latent and observed variables literally, because observed variables are not at all observed in the normal sense of the word. Second, the difference in handling observed and latent variables in actual data analysis cannot be defended by referring to the surplus meaning of latent variables as theoretical constructs, for observed variables carry such surplus meaning just as well. Moreover, in both cases it is necessary to make an inference from an observed data pattern to an underlying property. It seems that the distinction between latent and observed variables, which appears to be crucial in distinguishing latent variable modeling techniques from statistical techniques in general, has to be justified differently.* » (p. 30). Pour Borsboom c'est la probabilité de la véracité de la mesure qui différencie "observe" et "latent". Il n'y a pas de différence ontologique mais épistémique : « *the probability of an erroneous conclusion on the basis of data is judged differently. That is, if sex is conceptualized as an observed variable, then the researcher assumes that if, say, 1 is observed in the column that codes for sex, it is certain that the corresponding person is male. For a latent variable this is not the case. Upon observing an IQ score of 120, the researcher does not assume that it is certain that the person has above average intelligence; the person in question may, for instance, have been fortunate in*

guessing the right answers. Thus, the inference to the person's property (i.e., having above average intelligence) is, at most, probable given the data but not certain. When we treat a variable as observed, we mean nothing more than that we assume that the location of a person on that variable can be inferred with certainty from the data. When we treat a variable as latent, we mean that the inference in question cannot be made with certainty. It is important to see that this formulates the distinction between latent and observed variables as a purely epistemological distinction. As such, this distinction is partly a function of the observer. Thus variables are not inherently latent or observed. They can only be latent or observed with respect to the data at hand, or, in other words, with respect to the observer and his or her measurement equipment. There is no need, however, to assume that the distinction between latent and observed variables is an ontological distinction between different kinds of variables » (p. 31).

C'est en effet la dimension épistémique, sa connaissance, qui oblige à appréhender une propriété mentale comme un objet réel inobservable via une modélisation par variables latentes basées sur les manifestations de la propriété mentale. La dimension épistémique ne peut être abordée par les processus en amont, car elle est irréductible à eux épistémiquement ; mais de même cette propriété n'est pas réductible aux processus en interaction de ses manifestations. La propriété mentale doit être caractérisée épistémiquement par ses manifestations. Le recours à des variables latentes peut donc ouvrir un cadre opérationnel pour appréhender de telles propriétés émergentes.

1.1.6.2 Conclusion sur les variables latentes et l'ontologie des propriétés mentales

Les variables latentes ne seraient en définitive qu'un moyen d'opérationnaliser une réduction du complexe, sans contredire le niveau micro (réseau neuronaux), mais en donnant toute légitimité au niveau macro (les propriétés mentales) car seul niveau de compréhension potentiel. La démarche dès lors de « mesurer » une propriété mentale à partir de ses manifestations conserve toute sa légitimité pour inférer un état stabilisé du système complexe. Cette propriété mentale est « réelle », mais sans chercher à caractériser cette réalité au niveau neuronal (neuro-imagerie) ni par la construction d'un réseau complexe de ses manifestations (*Network Analysis*). La modélisation du complexe avec des variables latentes ne doit donc pas essayer de lier les réseaux de neurones (niveau micro) aux modèles avec variables latentes (niveau macro) comme ont tenté de le faire un certain nombre d'auteurs (Bagozzi, 2011a;

Charlton et al., 2008; Dietvorst et al., 2009; Kievit et al., 2011a, 2012), mais considérer que les variables latentes intègrent la complexité par un processus nécessaire de régression de cette complexité. De même, cette propriété émergente ne peut se réduire à la construction d'un réseau complexe de ses manifestations en oubliant la propriété émergente.

Avec les variables latentes, nous n'avons pas une vision modulaire comme le dit McGrath (2005a), nous avons une régression de la complexité de processus dynamiques non-linéaires. Les variables latentes deviennent ainsi non seulement légitimes, mais nécessaires car les propriétés émergentes ne peuvent se conceptualiser et être analysées qu'au niveau macro à partir de leurs manifestations ; sans pour autant remettre en cause le niveau micro comme niveau explicatif, et sans nécessité de faire appel à une émergence forte de ces propriétés comme le font certains auteurs (L. F. Barrett, 2011).

Il faut considérer que les variables latentes, formalisant les propriétés mentales, intègrent la complexité par un processus nécessaire de régression de cette complexité, tant conceptuellement que méthodologiquement. Avec une telle régression, il y a une perte irréversible d'information (pas d'analyse de la complexité en amont de la propriété mentale), et par ailleurs un gain d'information par une clarification que le niveau micro n'apporte pas (régression de la complexité), que perdent à mon sens les *Network Analysis*. Les propriétés mentales sont donc à considérer comme des réalités, en tant que attracteurs robustes d'un réseau dynamique complexe (les processus cognitifs de bas niveaux). Il est impossible de les définir plus précisément d'une façon spatio-temporelle comme le demande Maraun (2010) car pour une même tâche, une personne peut utiliser différents processus cognitifs (Berthoz, 2013)... mais ces propriétés mentales existent spatio-temporellement. Leur formalisation par des variables latentes devient dès lors difficilement contournable... sauf à les oublier comme les *Network Analysis*.

Mais comprendre une propriété mentale comme une qualité émergente d'un processus complexe, non réductible à ce processus, environnement dépendant et ayant des qualités à un niveau supérieur que le processus le générant, va obliger la psychologie à conceptualiser ces propriétés mentales dans une démarche particulière. Cela renvoie à la problématique des concepts en psychologie que je discute dans la section suivante, et en conséquence du cadre épistémologique nécessaire.

1.2 Le cadre épistémologique

Les propriétés mentales émergentes d'un réseau dynamique complexe sont des qualités non réductibles aux processus la générant, et exige un cadre conceptuel différent que ceux développés en neurobiologie. Cette conceptualisation au niveau du phénomène émergent oblige à discuter des concepts en psychologie, et de leur nature « objective » vs « subjective » ; et donc à discuter du cadre épistémologique nécessaire. Je reviendrai ultérieurement sur ce problème dans la partie qui développera le cadre que je considère valide de mesure des propriétés mentales ; cette partie ne fait que poser ici les bases de ce qui me paraît nécessaire comme cadre épistémologique en psychologie.

1.2.1 Le problème de la conceptualisation des propriétés mentales

Popper écrivait « *La psychologie est une science sociale car notre pensée et notre action dépendent dans une large mesure de conditions sociales [...] La psychologie présuppose donc des concepts sociaux.* » (Popper, 1979, p. 87). Cela revoie-t-il à une vision relativiste/construite des concepts en psychologie ? Les concepts caractérisant des propriétés mentales ne sont-ils que des construits sociaux ayant une fonction heuristique ?

1.2.1.1 Peut-on conceptualiser le phénomène émergent ?

La formalisation conceptuelle d'une propriété mentale est liée à ses manifestations, donc dépendante d'un cadre social, culturel, scientifique donné. L'évolution du langage employé en science est inscrite dans l'évolution du langage dans sa dimension pratique et opératoire en science. Il n'y en effet a pas d'invariance de signification¹⁹ (Feyerabend, 1979/2005). Mais ceci pourrait être énoncé pour tous les champs scientifiques.

Slaney & Racine (2013b) rappellent qu'il est commun de ne pas dissocier un concept (comme « une pomme ») telle que définit conceptuellement (une définition dans le champ de la botanique de ce qu'est théoriquement un pomme) et des objets que ce concept représente (des vraies « pommes » que nous dégustons). Le terme « pomme » renvoyant aux deux, le concept

¹⁹ L'incommensurabilité est « *la thèse selon laquelle les termes utilisés par une autre culture ne peuvent être équivalents par leur sens ou par leur référence aux termes ou aux expressions que nous utilisons aujourd'hui* » (Putnam, 1984). Je rejoins Putnam (Putnam, 1984) dans sa critique de l'incommensurabilité telle que posée par Kuhn (Kuhn, 1970) ou Feyerabend (Feyerabend, 1975) dans le sens où il y a une compréhension potentielle actuelle de textes passés (donc il n'y a pas incommensurabilité au sens strict) ; mais ceci n'est pas à mon sens contradictoire avec la non invariance des termes tel que Feyerabend le pose.

et l'objet réel. Mais pour ces auteurs, cette démarche est délicate (« *trickier* ») pour des concepts théoriques complexes comme les propriétés mentales. Maraun & Peters (2005) critiquent ainsi les concepts en psychologie car inscrits dans un vocabulaire non-épuré, mais socialement référencé, à la différence de l'écriture mathématique qui s'épure en grande partie de l'inscription socio-historique des termes employés (ceci est bien évidemment à relativiser, cf. par exemple la discussion de Hacking (2002) sur l'émergence des probabilités).

Les concepts psychologiques sont en effet des concepts complexes à définir formellement, ils ne sont pas définissables en termes techniques comme en science « dure » (comme la physique), et doivent avoir recours à un vocabulaire familier, avec donc toute la confusion possible entre un terme familier et un terme conceptualisé. C'est là un des problèmes essentiels pour certains auteurs, Lovasz & Slaney (2013) expliquent que le sens d'un concept en psychologie est inexprimable et flou et difficile d'articuler en des termes exacts. Maraun, Slaney, & Gabriel (2005) citent des exemples dans le *Diagnostic and Statistical Manual of Mental Disorders* (DSM) où par exemple la Schizophrénie n'est pas clairement définie. Pour eux la plupart des concepts psychologiques (considérés ici comme "*constructs*") ne sont pas définissables : « *it follows that constructs are indefinables and, hence, cannot be defined by necessary and sufficient conditions;* » et ils concluent : « *The true meaning of the concept-label is a pure essence that lies somewhere "behind" the phenomena.* » (p. 152).

1.2.1.2 Des concepts objectifs et intersubjectifs

Je vais amorcer la discussion à partir de L. F. Barrett (2009a) qui s'appuie entre autres sur Searle (1996). L. F. Barrett (2009a) revient sur le problème des concepts en psychologie à partir de la manière dont Searle (1996) propose de définir les concepts sociaux. Les scientifiques catégorisent avec des mots la réalité perçue. Les concepts catégorisant la réalité sont liés aux objectifs du scientifique et à ce que ce scientifique a comme cadre de référence (langage, culture, ...). La catégorisation en concepts de la réalité est donc dépendante de la démarche scientifique. Il y a des faits « objectifs » qui existent en dehors de notre conceptualisation, même si les concepts utilisés peuvent être fallacieux, ce sont des faits observateurs-indépendants. La démarche scientifique progresse et permet de mieux appréhender cette réalité. Il y a d'autre part des réalités subjectives, dépendantes de l'homme, comme en sciences sociales, qui existent par un consensus social (L. F. Barrett, 2009a, reprend les exemple de Searle : la monnaie, la nationalité, ...). La psychologie est en tension entre les deux : catégories objectives et catégories subjectives tel que Searle le propose.

L'émotion par exemple ne présente pas de fiabilité statistique à partir de relevés physiologiques (mesures cardiovasculaires, muscles faciaux, ... voir L. F. Barrett, 2009a). L'émotion n'existe pas uniquement dans la tête de l'observateur, il y a une réalité matérielle, mais l'émotion n'existe pas sans observateur, c'est-à-dire sans une interprétation subjective. Ces catégories subjectives existent par consensus social lié à un cadre, ils sont observateurs-dépendants, certains stables dans différentes cultures, certains très dépendants de la culture (p.e. l'émotion). Les catégories viennent de l'homme et constituent son expérience. L'état mental est classifié en fonction de sa fonction sociale de communication/régulation. Ces catégories instanciées sont réelles, mais dérivées par l'expérience humaine. Cependant les mots utilisés réifient les catégories, ils entraînent une démarche essentialiste des catégories alors que ce sont des concepts construits par l'homme.

Cette conceptualisation des états mentaux revient à considérer ces derniers comme des réalités intersubjectivées. L. F. Barrett (2009a) considère que puisque l'humain catégorise des objets, il y a une vision/réalité subjective de ces objets : « *The scientific ontology proposed here [...] deals with the existence of two domains of reality (one that is subjective and one that is objective) and their relation to one another.* » (p. 333). Les propriétés mentales que nous attribuons aux autres existent donc comme des fonctions sociales, et donc par consensus social lié à un cadre.

1.2.1.3 Des concepts stables et mouvants

Une propriété mentale conceptualisée est donc à la fois une réalité et une construction. La conceptualisation se fait sur la base d'expériences réelles, donc la démarche est réaliste avant d'être subjective (Madelrieux, 2012). Autrement dit, le processus de connaissance ne s'enferme pas dans le seul univers des représentations (le langage entre autres) : il y a systématiquement confrontation à l'expérience du réel (Avenier, 2011). Le langage qui formalise ces expériences du réel n'est pas un objet en soi, pas une entité séparée du monde, c'est un outil d'interprétation du monde, il est dialectiquement lié aux expériences, donc au réel.

Si une personne n'est jamais dans un état stable, car toute personne est en évolution permanente (Andrieu, 2007), on perçoit cependant des stabilités psychologiques chez autrui, et les relations sociales se construisent à partir de cette stabilité psychologique rationalisée (Ludwig, 2004). Cramer et al. (2012) considèrent « *For some reason, human systems tend to settle in relatively fixed areas of the enormous behavioural space at their disposal, where they*

are in relative 'equilibrium' with themselves and their environments. By equilibrium, we mean a stable state in this definition ». Les propriétés mentales sont réifiées (L. F. Barrett, 2009a), permettant les relations sociales sur la base de catégories. Mais les concepts ne sont pas que des construits sociaux un peu artificiels, ils représentent une réalité. Hacking (2008) « *Construit social et réel semblent terriblement s'exclure l'un l'autre. Une partie de la tension entre les deux provient de l'interaction entre les deux* » (p. 141) et cite Putnam : « *une erreur philosophique commune quand on suppose que la "réalité" doit faire référence à une seule superchose, au lieu de faire attention aux manières que nous avons de renégocier en permanence – et nous sommes obligés de renégocier – notre notion de réalité au fur et à mesure que notre langue et notre vie se développent* ».

Par ailleurs, les propriétés mentales réifiées deviennent des objets réels par ce que Hacking appelle « *l'effet de boucle* ». La propriété mentale réifiée entraîne des modifications chez la personne et son entourage social et donc la réification sociale génère en boucle cet objet réifié comme une réalité. Pour exemple, caractériser une personne de dépressive entraîne chez cette personne et son entourage une réalité de la « *dépression* » même si au départ ce n'est qu'une conceptualisation/réification à partir de manifestations diverses. Les objets étudiés en psychologie sont en effet des objets ayant une fonction sociale, les propriétés mentales sont la base de l'existence sociale des individus : « *le vocabulaire des concepts sociaux permet de décrire des états internes de nos semblables, et par là d'expliquer leur comportement observable en les rationalisant* » (Ludwig, 2004, p. 37), et donc la perception de ces propriétés mentales par le cadre social influence en retour ces propriétés mentales : « *les idées interagissent souvent avec des états, des conditions, des comportements, des actions et des individus [...] l'interaction intervient dans la matrice plus large des institutions et des pratiques qui entourent une classification* » (Ludwig, 2004, p. 37).

Cette réification entraîne par cet effet de boucle un *mouvement* des propriétés mentales. (Hacking, 2008, pp. 143–145) « *Il y a une tendance constante dans les sciences sociales et psychologiques à imiter les sciences naturelles et à produire de vrais genres naturels de personne [...] Il existe une représentation de l'objet à fouiller pour aboutir au genre exact, le genre qui est conforme à la nature, une cible fixe pour peu qu'on puisse l'atteindre. Mais peut-être s'agit-il d'une cible mouvante simplement à cause de l'effet de boucle associé aux genres humains ? C'est-à-dire que de nouvelles connaissances transforment la manière dont ces individus se comportent et reviennent en boucle pour provoquer des changements dans la classification et dans la connaissance que l'on en a* » (p. 147). Pour Hacking (2008) « *les cibles des sciences naturelles sont stationnaires [...] les cibles des sciences sociales sont en*

mouvement » (p. 151). On retrouve aussi cette vision de concepts formalisés qui deviennent des objets réels car socialement existant et mouvants chez Desrosières (2008a, 2010) au sujet des indicateurs économiques quantifiés (qui sont eux aussi des constructions sociales mais qui deviennent des réalités quand ils servent de référents aux orientations politiques et économiques, comme le PIB par exemple).

Les propriétés mentales ne sont donc pas que des « construits » au sens de « construction de l’imaginaire du chercheur », ce qui interdirait sa généralisation ; mais elles ne sont pas non plus qu’une réalité objective inerte, c’est-à-dire entièrement indépendantes de l’intersubjectivité. Les propriétés mentales sont des objectivités-intersubjectivées, stables et mouvantes, qui doivent dès lors être comprises comme des « *réalités en mouvement* » (Hacking, 2008), autant objectives qu’intersubjectives, et que j’appellerai “concepts mous” car se référant à un objet ayant une certaine consistance (attracteur d’un réseau ayant une robustesse) mais modifiable au gré des interactions sociales.

1.2.1.4 Signifiant vs signifié

La critique de Maraun et al. (2005) sur le flou de la conceptualisation des concepts en psychologie, notamment chez Cronbach & Meehl (1955), me paraît discutable. Ces auteurs illustrent le flou scientifique de Cronbach et Meehl par les mots de Cronbach & Meehl (1955, p. 294) : « *We will be able to say ‘what anxiety is’ when we know all the laws involving it; meanwhile, since we are in the process of discovering these laws, we do not yet know precisely what anxiety is* ». Cronbach et Meehl disent en effet que l’on ne sait pas ce qu’est l’anxiété, c’est-à-dire on ne sait pas le définir conceptuellement de façon rigoureuse, mais on sait qu’il y a un phénomène qui provoque des données observables ayant une cohérence, et donc on cherche à mieux cerner ce phénomène. L’anxiété n’est pas pure « essence » ici, lorsque Cronbach et Meehl disent « *on ne sait pas ce qu’est précisément l’anxiété* » cela renvoie à la complexité de conceptualiser un phénomène inobservable dont on peut supposer qu’il est à l’origine de manifestations observables récurrentes.

La critique de Maraun Slaney et Gabriel semble revenir à appliquer ce qu’ils reprochent aux autres : assimiler le phénomène au concept. Pour eux, puisque le concept est parfois mal défini, cela revient à considérer que le phénomène est une pure « essence » sans autre consistance pour le théoricien. Or, une propriété mentale est définie verbalement, et cela limite la capacité de définir de façon exhaustive et claire le concept censé caractériser cette propriété mentale. Mais cette difficulté de conceptualiser un phénomène déborde largement la

psychologie car cette difficulté est inhérente à tous les champs scientifiques. Soler (2009) discute de cela d'une façon plus générale : *« il est fort difficile de donner un sens précis à l'idée même d'une correspondance entre une entité linguistique (énoncé ou un ensemble d'énoncés) et un référent extralinguistique : en quoi la signification coordonnée à une suite de mots peut être dite semblable à des états de choses réels ? C'est toute la question, complexe et hautement polémique, de la nature des rapports entre signification et référent qui se trouve en jeu »* (p. 156).

1.2.1.5 Conclusion

Les propriétés mentales, qui émergent d'un réseau dynamique complexe, sont des propriétés non réductibles aux processus le générant et exige un cadre conceptuel différent de la neuro-imagerie. Cette conceptualisation ne peut se faire que sur la base des manifestations de ces propriétés mentales qui sont non-réductibles aux processus qui les ont créées. Le cadre conceptuel d'une propriété mentale est en tension entre une réalité observable (les manifestations concrètes de la propriété émergente) et la subjectivisation de ces qualités car toujours inscrites dans un cadre social. Un concept renvoie à une réalité (le terme théorique), mais dans le cadre des concepts en psychologie, cette réalité ne peut s'appréhender que par des manifestations, donc dans un cadre social. Une propriété mentale conceptualisée permet de concilier un terme théorique (un concept) et le référent réel de ce concept (la propriété mentale émergente observée par ses manifestations). La propriété mentale n'est donc pas qu'un « construit » au sens de « construction de l'imaginaire du chercheur », ce qui interdirait sa généralisation ; mais en revanche elle n'est pas qu'une réalité objective externe à la recherche et à la vie sociale. Cette tension entre réalité et subjectivité pousse la psychologie à devoir renégocier en permanence la propriété mentale (comme le dit Putnam cité précédemment).

1.2.2 Rupture épistémologique avec l'empirisme-réaliste

La formalisation des propriétés mentales renvoie à une démarche scientifique en tension entre le réalisme et le constructivisme. Le caractère relativiste de nos connaissances ne s'oppose pas à l'objectivité au sens où les propriétés mentales sont des réalité externes à l'observateur, en considérant que cette objectivité est une objectivité « pratique », dans un sens de « pratique sociale » (observateurs-dépendants), donc renégociable. La démarche

épistémologique nécessaire doit donc considérer les propriétés mentales comme construites dans un rapport dialectique entre une réalité et une expérience pratique du scientifique.

1.2.2.1 Réalisme et constructivisme

La formalisation d'une propriété mentale est subjective, donc dépendante d'un cadre social, culturel, scientifique donné. Cela revoie-t-il à une vision relativiste/construite des concepts en psychologie ? La subjectivité est liée à toute démarche scientifique qui verbalise une réalité non-verbale, et cette démarche n'est pas contradictoire pour moi avec un positionnement épistémologique réaliste. La psychologie a en effet, comme tout champ scientifique, une démarche pour clarifier les concepts (comme l'émotion) en liant la théorie à l'empirisme, mais c'est bien l'observation de la réalité qui modifie les concepts, ils sont donc bien réalité-dépendants. Les propriétés mentales ont ainsi une réalité matérielle sous-jacente à leur conceptualisation (ou du moins, cela est posé comme hypothèse et la démarche scientifique cherchera à valider cela).

Dire qu'une propriété mentale est « réelle » n'équivaut pas nécessairement comme le disent Slaney & Racine (2013b) à être capable de l'observer en tant qu'objet, c'est-à-dire pouvoir la définir spatio/temporellement. En effet, les propriétés mentales de par leur complexité et de par leur caractère émergent vis-à-vis des réseaux neuronaux peuvent être considérées à la fois comme « réelles » tout en étant non-réductibles à un état neuronale précis généralisable. Etant dynamique, modifiable, et en permanence soumis à des stimuli sensoriels externes, socialement inscrits, le réseau de neurone peut trouver des états stationnaires (propriétés mentales) non-réductibles à un état neuronal précis (un système émergent est indépendant de variations minimales des éléments microscopiques). La neuro-imagerie n'est donc certainement pas à même d'observer ni de caractériser ces propriétés mentales de façon satisfaisante pour les raisons évoquées précédemment. Ceci ne retire en rien la matérialité de ces propriétés mentales comme état stationnaire (attracteur du système non linéaire complexe que représente le réseau neuronal), mais cette matérialité ne peut se caractériser que par les manifestations concrètes. Leur catégorisation par des concepts les classant en catégories différentes est certes un artifice humain, comme toute classification. Cette réification permet de construire des théories où des processus mentaux suffisamment « proches » (au sens où leurs manifestations le sont) peuvent être caractérisés par une seule et même catégorie linguistique. En effet, si les propriétés mentales ne sont pas inertes, on retrouve suffisamment de stabilité pour certaines

d'entre elles pour que la pratique sociale les réifie/catégorise... et renégocie cette catégorisation.

1.2.2.2 La nécessité d'un cadre épistémologique particulier

Le cadre épistémologique nécessaire comme théorie de la connaissance doit donc concilier la dimension réaliste des propriétés mentales, mais réalité mouvante ; et leur conceptualisation qui catégorise des propriétés ayant une stabilité suffisante (au sens où la pratique sociale renvoie une certaine stabilité chez une personne pour une certaine propriété mentale). Ce cadre épistémologique doit clairement affirmer son caractère « pratique » au sens de l'intersubjectivité et de la renégociation permanente nécessaire de ces catégories. Il faut par nécessité, lorsque l'on travaille sur des propriétés mentales, poser un cadre épistémologique qui renvoie à la fois au réalisme et au constructivisme.

On ne peut donc pas considérer, comme l'empirisme-réaliste le pose, qu'il y a des « réalités » indépendantes du chercheur lorsque l'on travaille sur des propriétés mentales, c'est-à-dire des objets que nous pourrions définir précisément comme dans les sciences de la nature. A l'inverse, on ne peut pas considérer ces propriétés mentales uniquement comme des « constructions » sans soubassements réels. Il faut donc une épistémologie en rupture avec l'empirisme-réaliste qui recherche le « vrai » comme une entité en soi ; sans entrer pour autant dans un constructivisme radical car les propriétés mentales existent indépendamment des représentations que nous en avons. Cette réalité ne peut être appréhendée indépendamment de la pratique qui en est faite, nous avons qu'une connaissance pratique des propriétés mentales qui par ailleurs sont des objets mouvants du fait de cette pratique sociale.

1.2.2.3 Le recours au pragmatisme

Il y a donc une tension entre un « réalisme » et un « constructivisme » dans l'usage des concepts en psychologie, que certains auteurs catégorisent en « constructivisme réaliste » (Markus & Borsboom, 2013a; Mislavy, 2009). Certains auteurs en psychologie se réclament de ce courant (Lektorskii, 2010; Mislavy, 2009; Noaparast & Khosravi, 2006), mais le cadre formel de celui-ci semble encore peu défini en psychologie et plus usuel en sociologie. En fait, en psychologie, beaucoup d'auteurs renvoient leur démarche plutôt au « pragmatisme » pour caractériser l'aspect opératoire d'une démarche scientifique liant une forme de réalisme et une forme de constructivisme. Nous pouvons en effet considérer que le cadre épistémologique nécessaire aux travaux avec des propriétés mentales doit s'inscrire dans le

pragmatisme. Mais il est toujours délicat de se référer à un courant affirmé sans risque et il me paraît préférable d'explicitier ce cadre épistémologique qui me paraît nécessaire.

Le pragmatisme, au sens philosophie des sciences, est généralement relié à James, Peirce et Dewey, sans pour autant être un véritable courant philosophique homogène ; *« ce n'est pas "une théorie systématique", c'est plutôt "un certain groupe de thèse" »* (Tiercelin, 2002). La problématisation qui relie ces auteurs est une démarche scientifique en tension entre le réalisme et le constructivisme. Pereira (2012) considère que *« le pragmatisme ne nie pas forcément la correspondance de la vérité et de la réalité, mais il insère aussi bien la conception correspondantiste et la conception cohérentiste dans une conception plus large à la fois naturaliste et sociale de la vérité [...] dans le pragmatisme, la théorie est toujours pensée à partir de la pratique [...] il n'y a pas de sens à séparer les deux dans une démarche pragmatique [...] refus de séparer théorie et pratique »*. Le pragmatisme intègre le fait que le scientifique a un projet dans sa démarche. Le pragmatisme est une « méthode » plus qu'une philosophie au sens strict. Il se démarque d'une conception statique de la raison, il privilégie les processus et démarches sans entrer dans un relativisme intégral. Le pragmatisme est en rupture avec l'empirisme-réaliste qui recherche le « vrai ». La science se construit sur des manifestations qui sont liées à ce qu'une expérience permet d'observer, donc il y a un effet conjoint de l'acteur et de l'objet étudié dans l'activité scientifique. Toute différence présumée doit se traduire dans une différence observable dans la pratique. *« il ne semble pas qu'il y ait une importante distinction entre deux propositions qui ne puisse se traduire dans les effets pratiques différentes [...] la signification intellectuelle de toutes pensées réside en dernière instance dans ses effets sur nos actions »* (Peirce, cité par Cometti, 2010). Ce caractère relativiste de nos connaissances ne s'oppose pas à l'objectivité en considérant cette objectivité comme « pratique », et c'est ce caractère de praticité de la théorie qui la valide ou non ; donc toute théorie est contingente des objectifs et du langage, mais dépendante d'une réalité ; et donc en contradiction avec le relativisme radical.

Il est à noter que la vision « pratique » du pragmatisme est souvent vue dans un sens étroit, c'est-à-dire avec un sens purement applicatif. Pour moi, la « pratique » associée au pragmatisme n'a pas qu'une vision réductrice, dans le sens « opérationnalisation » de l'outil (donc instrumentaliste) comme Markus & Borsboom (2013a) le considèrent. Le sens « pratique » doit être pris dans un sens de « pratique sociale », c'est-à-dire incluant dans la pratique les dimensions conscientes et inconscientes d'un chercheur prisonnier du cadre psychosociologique de sa fonction et de son époque. Le terme *praxis* est donc peut être plus opératoire et je l'utiliserai parfois.

Le discours scientifique en psychologie est pragmatique car basé sur une praxis, il est construit dans un rapport dialectique entre la réalité extérieure au scientifique et l'expérience pratique du scientifique. La « vérité » n'existe pas « en soi ». Par ailleurs, le pragmatisme rejoint la démarche de Popper sur la falsifiabilité et le caractère faillible de nos connaissances, Dewey parlait de *assertibilité garantie*, car toute différence présumée entre deux modèles doit se traduire dans une différence observable dans la pratique (Cometti, 2010).

1.2.2.4 Pragmatisme et pragmatisme

La littérature académique utilise le terme « pragmatisme » sans toujours clairement différencier le terme commun et le terme renvoyant au courant de la philosophie des sciences. Markus & Borsboom (2013a) par exemple se démarquent clairement du « pragmatisme » au sens philosophique au début de leur ouvrage, mais usent ensuite à plusieurs reprises du terme « approche pragmatique » pour caractériser ce qui leur semble devoir être fait. Dans la littérature, le terme « pragmatisme » prend ainsi souvent le sens de « orientation pratique », avec une connotation « instrumentaliste ». Pour exemple : « *Très souvent les sciences sociales, mais aussi la psychologie, semblent se trouver dans la situation où c'est précisément la mesure qui crée le concept. Cet aspect pragmatique est la conséquence de la complexité inhérente aux objets de ces domaines, mais aussi de la portée sociale et politique que peuvent y avoir les résultats des recherches.* » (Volken, 2007) ; Desrosières (2008a) : « *La composante pragmatique de la mesure dans la méthodologie des sciences sociales s'illustre clairement dans le fait que les attributs mesurés ne sont finalement définis, la plupart du temps, que par le processus de la mesure.* » ; ou Markus & Borsboom (2013a) qui différencie le constructivisme-réaliste de Messick du pragmatisme de Kane car ce dernier oriente sa vision des tests vers le côté pratique et fonctionnel en rupture avec le réalisme : « *Kane's approach to validation reflects shift from realism toward pragmatism in the following sense [...] the interpretative argument as the basis for the validity argument [...] more closely tied to communication and dialog [...] context-specific argument constructed at a particular time and place for a particular purpose [...] depend upon the specific interpretation at hand [...] validation in context-specific features* » (p 12-13).

Clairement, le pragmatisme auquel je me réfère n'est pas ce pragmatisme avec un petit « p » (comme le pose Kane (2013b) explicitement). Je peux caractériser la démarche pragmatique à laquelle je me réfère, avec un grand P, comme Pereira (2012) le définit : « *le pragmatisme, au sens strict, désigne un courant de la philosophie américaine divisé en deux sous courants : le*

pragmatisme classique (dont les principaux représentants sont Pierce, James et Dewey) et le néo-pragmatisme principalement incarné par Putnam, Rorty et Shusterman. Mais on peut considérer que le pragmatisme au sens large incarne une position structurelle de la philosophie [...] Le pragmatisme se caractérise par un refus de toute démarche de recherche d'un principe premier, d'un fondement [...] Donc en premier élément, d'un point de vue pragmatiste, il apparaît tout d'abord que toute réalité est relative au point de vue singulier qui est adopté sur la réalité et nous n'avons accès, par l'intermédiaire du langage, qu'à des points de vue humains. ».

1.2.2.5 Pragmatisme et constructivisme

Le pragmatisme peut sembler proche du constructivisme-réaliste (Fine, 2001) ; et parfois ces deux courants sont assimilés. Le constructivisme fonde son étude de la science sur la recherche « en acte » et sur les pratiques quotidiennes des scientifiques qui développent la science (Schinckus, 2008). Le pragmatisme et le constructivisme intègrent l'intentionnalité du sujet connaissant et récusent la connaissance objective indépendante du sujet. Ainsi on peut trouver un pragmatisme proche du relativisme qui « *refuse l'existence de choses qui ne soient pas relationnelles* » comme chez Rorty (Cometti, 2001), et laissant le primat de la connaissance au langage et ses limites (Chauviré, 2010). Ce pragmatisme rejoint le constructivisme (radical) qui met le primat sur l'individu et donc « *un renoncement à la valeur de vérité objective* » (Le Moigne, 2010, p. 68). « *Le constructivisme supporte trois propositions : (1) la pratique scientifique est "opportuniste", (2) la pratique scientifique n'a de réalité que locale et (3) les faits scientifiques sont engendrés par les détails contingents de la pratique scientifique. Ne disposant d'aucun critère extralinguistique, nous n'avons aucun critère de démarcation qui nous permettrait de distinguer ce qui viendrait de notre langage de ce qui viendrait de la réalité elle-même.* » Ainsi pour ces courants, « *l'objectivité d'un fait scientifique n'est pas dans sa préexistence à la démarche scientifique, mais bien dans la manière de construire ce fait scientifique au sein même de la démarche scientifique* » (Schinckus, 2008).

On peut contester cette vision du pragmatisme-relativiste ou constructivisme-radical. Appréhender la réalité comme subjectivée n'est pas en contradiction avec une position « réaliste », c'est-à-dire considérant qu'il y a des objets réels externes à notre conscience, même si le processus d'objectivation est interne à notre conscience. Une forme de pragmatisme adhère à cette critique du constructivisme radical en refusant de nier la prise en

compte des réalités objectives. L'expérience est pratique et active dans le pragmatisme de Dewey (1925/2012), il y a une interaction du sujet et de son environnement ; l'expérience est donc réaliste avant d'être subjective (Madelrieux, 2012). La philosophie pragmatique est une « *philosophie des flux* » entre l'objet et son objectivation (Foucart, 2014; Pereira, 2009). L'expérience, l'expérimentation, sont toujours incomplètes ; la réalité est ouverte et inachevée, de même que la connaissance. Ce pragmatisme se détache du constructivisme radical en affirmant que le monde existe indépendamment des représentations que nous en avons, même si la connaissance se fait par une praxis ; l'objectivation du monde est autant réelle que construite.

La différence entre une forme de constructivisme et une forme de pragmatisme semble parfois faible. Si nous reprenons le livre de Le Moigne (2010) définissant le constructivisme, je peux adhérer dans les grandes lignes à sa démarche fondamentale. Pour Le Moigne, la connaissance n'est plus démontrée, mais elle est « argumentée », donc constructible et reproductible, de façon à permettre leur intelligibilité pour son interlocuteur préalablement informé des hypothèses épistémologiques qui légitiment cette argumentation dialectique et tâtonnante. Pour conserver l'externalité de la valeur de la connaissance, on parle de « vérité intersubjective » qui implique un pragmatisme. Ce courant constructiviste accepte l'idée d'une externalité de l'objet de notre connaissance... et se rapproche alors du pragmatisme dans sa méthode. Pour Le Moigne (2010), il y a « *un renoncement à la valeur de vérité objective [...] pour substituer la vérité objective à la vérité révélée* » (p. 68). Ainsi « *la connaissance humaine est processus avant d'être résultat, elle se forme dans l'action et l'interaction* » et pour lui « *L'hypothèse gnoséologique fondatrice des épistémologies constructivistes est phénoménologique : la connaissance humaine est processus avant d'être résultat, elle se forme dans l'action et l'interaction [...] Elle est projective, plutôt que subjective ("la médiation de l'objet par le sujet prend toujours la forme du projet" Bachelard, p. 117-118).* ».

La tension entre constructivisme et pragmatisme est à mes yeux sur « l'objectivité de la connaissance » à travers le langage, c'est-à-dire que le pragmatisme est en rupture avec le scepticisme du constructivisme. Le Moigne (2010) considère « *un renoncement à la valeur de vérité objective [...] pour substituer la vérité objective à la vérité révélée* ». Ce constructivisme tend vers un relativisme de la connaissance que critique le pragmatisme. D'une manière à mon sens plus radical, Schinckus (2008) considère la connaissance de la façon suivante « *Ne disposant d'aucun critère extralinguistique, nous n'avons aucun critère de démarcation qui nous permettrait de distinguer ce qui viendrait de notre langage de ce qui*

viendrait de la réalité elle-même [...] la relation de référence entre les objets et les mots est relative à un choix arbitraire de schème de référence (ou à un manuel particulier de traduction) [...] la vérité, la référence et l'ontologie sont relatives à une théorie d'arrière-plan ou à un langage particulier et deviennent immanentes à ce langage ou à cette théorie. ». C'est peut-être cette démarcation entre réalité et langage propre au constructivisme (et relativisme) que cherche à rompre le pragmatisme.

Le pragmatisme ne nie pas l'objectivité de l'objet de la connaissance, même si c'est une objectivité pratique. Car si, comme le dit Cometti (2001) « *Le pragmatisme refuse d'admettre l'existence de choses qui ne soient pas relationnelles* » (donc en phase avec le constructivisme), « *Le pragmatisme substitue à l'idée d'une rationalité fixe et close celle d'une rationalité ouverte qui renvoie dos à dos relativisme et anti-relativisme, et qui échappe au scepticisme moral en se refusant à admettre que l'absence de justification ou de fondement équivaut à l'arbitraire des choix.* » car « *l'expérience n'est pas d'abord cognitive, mais pratique et active (c'est l'interaction de l'organisme et de son environnement) ; l'expérience n'est pas d'abord subjective, mais réaliste (c'est la nature qui est expérimentée et qui répond à nos actions).* » Le langage n'est pas à considérer comme indépendant de la réalité « *Quant à l'entreprise de faire se détacher le monde sur fond de langage, elle est vaine puisque monde et langage sont co-extensifs comme l'envers et l'endroit d'une même réalité* » (Chauviré, 2010) ; ou comme le dit Avenier (2011) « *l'hypothèse d'inséparabilité (dans le processus de connaissance) entre le système observant et le système observé* ». Ainsi pour Madelrieux (2012) « *l'expérience humaine est une partie de la nature, si bien que, encore une fois, la connaissance ou la subjectivité (mais aussi les significations, la conscience, l'art, les valeurs, etc.), qui sont des tard-venus dans l'évolution, doivent être définis en termes d'interaction naturelle et non en termes d'entité transcendantale – la relation entre sujet et objet de connaissance n'étant ainsi qu'une complexification et une spécialisation de l'interaction entre organisme et environnement.* » et précise en parlant du pragmatisme « *C'est l'occasion d'opposer son naturalisme à toute forme de dualisme, puisque cette continuité de la nature montre qu'il n'existe pas deux mondes séparés, l'un physique et l'autre psychique, mais un seul et même monde possédant des degrés de complexité variés, depuis l'interaction physico-chimique jusqu'à la communication humaine* ». C'est là à mon sens la fracture avec le constructivisme qui questionne la « réalité » de notre intelligibilité. Le Moigne (2001) considère que la thèse de Searle (« *le réalisme comme condition d'arrière-plan de toute intelligibilité* ») est questionnable car « *rien ne m'oblige à en conclure définitivement que cette hypothèse est la seule qui soit plausible ni a fortiori la seule qui soit universellement*

évidente ». Mais j’approfondirai ce point lorsque je développerai le cadre pratique de la mesure.

1.2.2.6 Pragmatisme-réaliste

Comme le pragmatisme ne couvre pas un courant homogène (« *or le terme "pragmatisme" recouvre aujourd’hui des acceptations et des auteurs qui s’opposent point par point* », Tiercelin, 2002), pour se différencier d’un pragmatisme proche du relativisme ou du constructivisme radical, je parlerai de pragmatisme-réaliste. Tiercelin (2002) discute du « *réalisme pragmatique* », mais je préfère poser pragmatisme-réaliste car la dimension pragmatiste paraît première dans cette formulation. Le « pragmatisme-réaliste » est parfois assimilé au « constructivisme-réaliste » (Fine, 2001). Je préfère conserver le terme « pragmatisme » car il intègre une démarche de la science clairement inscrite dans une « praxis » au sens « social ». De même, la fonction du langage s’oriente vers une objectivité avec le pragmatisme et rompt avec une forme de relativisme propre au constructivisme. Nos formalisations via le langage sont intrinsèquement reliées à la réalité appréhendée de façon pratique, et non comme séparés de cette réalité. Ce pragmatisme permet d’éviter une sorte de cadre épistémologique voyant une science purement instrumentaliste, et donc relativiste. Le pragmatisme-réaliste se démarque ainsi du scepticisme de la connaissance porté par le courant constructiviste.

Le pragmatisme-réaliste ainsi caractérisé semble être une théorie de la connaissance à même de concilier la dimension réaliste des propriétés mentales et leur conceptualisation subjective, car inscrit dans une pratique scientifique concrète se confrontant à la réalité. Putnam est souvent associé à ce courant (Sosa, 1993). « *Putnam veut montrer [...] que nos conceptions, pour indissociables qu’elles soient de notre langage et de notre forme de vie, n’en sont pas moins objectives, même si leur objectivité est une « objectivité-pour-nous* » (Cometti, 2010, p. 132). Pour Putnam « *le fait de rejeter l’idée qu’il existe une perspective « externe » cohérente, une théorie qui serait simplement vraie en soi, indépendamment des observateurs possibles, ne revient pas à identifier la vérité à l’acceptabilité rationnelle [...] la vérité est une idéalisation de l’acceptabilité rationnelle, le terme « idéalisation » garanti une certaine stabilité, et distingue « vérité » et « acceptabilité rationnelle* ». Cette « objectivité pour nous » posé par Putnam est donc clairement en rupture avec le relativisme et le scepticisme.

S’il me paraît nécessaire de me décaler de Bourdieu en psychologie, car le surdéterminisme et le cloisonnement radical en « champs » de Bourdieu me paraît quelque peu critiquable, je

reprends à mon compte ses remarques critiques du réalisme et du relativisme pour le processus de connaissance en sciences sociales. Bourdieu écrit (2001, pp. 146-151) : « *On peut donc dire qu'il n'est pas de réalité objective indépendante des conditions de son observation sans mettre en doute le fait que ce qui se manifeste, une fois ces conditions déterminées, conserve un caractère d'objectivité [...]. Substituer un rapport entre un sujet (le savant) et un objet, un rapport entre les sujets (l'ensemble des agents engagés dans le champ) à propos du rapport entre le sujet (le savant) et son objet, conduit à rejeter à la fois et la vision réaliste naïve selon laquelle le discours scientifique est un reflet direct de la réalité, un pur enregistrement, et la vision constructiviste relativiste, selon laquelle le discours scientifique est le produit d'une construction, orientée par des intérêts et des structures cognitives, qui produirait des visions multiples, sous-déterminés par le monde, de ce monde [...]. La science est une construction qui fait émerger une découverte irréductible à la construction et aux conditions sociales qui l'ont rendue possible* ». De ce fait je rejoins aussi Bourdieu (2001, p. 147) dans sa critique de Feyerabend : « [...] *la lutte scientifique, à la différence de la lutte artistique, a pour enjeu le monopole de la représentation scientifiquement légitime du « réel » et que les chercheurs, dans leur confrontation, acceptent tacitement l'arbitrage du réel (tel qu'il peut être produit par l'équipement théorique et expérimental effectivement disponible au moment considéré) [...]. Tout se passe comme si [...] les chercheurs s'accordaient tacitement sur le projet de donner une représentation réaliste du réel ; ou plus précisément, acceptaient tacitement l'existence d'une réalité objective par le fait de se donner le projet de chercher et de dire la vérité du monde et d'accepter d'être critiqués, contredits, réfutés, au nom de la référence au réel, ainsi constitué en arbitre de la recherche [...]. Ce postulat ontologique en suppose un autre, l'idée qu'il y a du sens, de l'ordre, une logique, bref quelque chose à comprendre dans le monde, y compris dans le monde social [...]; que l'on ne peut pas dire n'importe quoi à propos du monde (« anything goes » selon la formule chère à Feyerabend), parce que tout est n'importe quoi n'est pas possible dans le monde* ».

1.2.2.7 Une démarche abductive

Le pragmatisme est généralement associé à l'abduction. Comme je me servirai de l'abduction par la suite, je pose les fondements de la démarche abductive et je renvoie à Haig pour un développement plus exhaustif. Haig a en effet écrit un certain nombre d'articles sur la méthode abductive nécessaire à ses yeux en psychologie, en donnant à mon sens un cadre un

peu normatif ; mais son travail permet de comprendre l'intérêt de cette approche (Haig, 2005a, 2005b, 2008b, 2008c).

Le cadre formel de la démarche scientifique reconnaît en général deux méthodes : l'induction et la déduction. L'abduction a été proposée comme méthode initiale à toute démarche scientifique par Peirce. Pour Peirce (1879/2014), c'est le doute issu d'une expérience contradictoire entre la réalité et les croyances scientifiques qui poussent à créer de nouvelles croyances scientifiques. Fidèle au pragmatisme, pour Peirce le chercheur n'est pas un réceptacle passif mais il est actif. L'esprit scientifique est une intelligence créative. L'abduction est une démarche qui n'est ni inductive ni déductive car la nouvelle règle posée n'est qu'une hypothèse. Cette hypothèse doit pouvoir être vérifiée expérimentalement. Cette hypothèse s'appuie sur les connaissances antérieures, mais représente aussi une nouveauté. Ce n'est donc ni une induction ni une déduction car ni l'une ni l'autre ne crée de nouvelle théorie, avec l'abduction il y a une « intelligibilité créatrice ».

Peirce donne pour illustration :

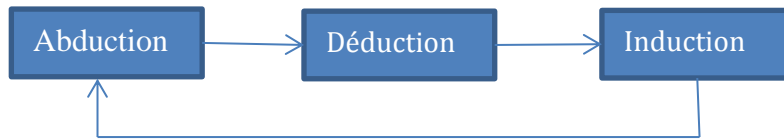
TABLEAU 1 : EXEMPLE SUR L'ABDUCTION

Déduction	A. Tous les haricots de ce sac sont blancs (règle) B. Ces haricots viennent du sac (cas) C. Ces haricots sont blancs (conséquence)
Induction	B. Ces haricots viennent du sac (cas) C. Ces haricots sont blancs (conséquence) A. Tous les haricots de ce sac sont blancs (règle)
Abduction	A. Tous les haricots de ce sac sont blancs (règle) C. Ces haricots sont blancs (conséquence) B. Ces haricots viennent du sac (cas)

Tiré de de Peirce (1879/2014)

La déduction (ABC) consiste à tirer une conséquence « logique » à partir d'une règle A et d'une observation B. L'induction consiste à trouver une règle A à partir de faits empiriques. L'abduction élabore une hypothèse qui cherche à relier la règle (les connaissances) et l'observation empirique. Pour Peirce, la séquence en science est :

FIGURE 11 : SEQUENCE DU PROCESSUS DE RECHERCHE



Tiré de Peirce (1879/2014)

Très clairement, l'abduction n'est pas à considérer comme de simples « *inferences to the best explanation* » (Michell, 2012, p. 258), c'est-à-dire dans un sens étroit. L'abduction permet d'hypothéser une nouvelle règle scientifique. La déduction consiste à inférer les conclusions de l'hypothèse. L'induction permet de vérifier la consistance de l'hypothèse en confrontant l'hypothèse à l'expérience. La confrontation permanente des théories et des faits entraîne la répétition de cette boucle. Nous retrouvons une démarche très poppérienne : « *La connaissance ne commence pas par des perceptions ou des observations, par une collection de données ou de faits, mais bien par des problèmes* » (Popper, 1979, p. 76). L'abduction renvoie ici à une rupture avec le logicisme formel, même si un certain nombre de publications semblent essayer d'intégrer l'abduction dans la logique formelle. C'est une manière de concevoir la démarche scientifique en tension entre objectivité logique et supputation subjective, sans forcément de ressort logique (l'intuition du chercheur) au cœur des découvertes scientifiques, mais plus fondamentalement d'une démarche pratique et concrète de la science qui ne ressemble pas au formalisme logicien (Feyerabend, 1979/2005).

L'abduction est intrinsèquement reliée à une démarche pragmatique car « la théorie » est une interaction subjective entre le chercheur et un processus social de consensus et de remise en cause de ce consensus (Kuhn, 1970). Lorsque je discuterai de la mesure, je reviendrai sur l'abduction car il y a avec les propriétés mentales une étape « abductive » où l'on remonte de résultats au niveau d'hypothèses, comme le dit Borsboom (2008b) : « *the inference that a set of indicators are affected by the same latent variable requires an abductive step: this means that the acceptance of the latent variable hypothesis is not mandated by the data alone, but requires an appeal to the explanatory merits of this hypothesis. Such explanatory merits may, for instance, involve strong theory about the processes that connect the latent variable to its indicators (Borsboom, Mellenbergh, & Van Heerden, 2004) and coherence of the hypothesis with a body of accepted knowledge (Haig, 2005a, 2005b).* ».

1.2.2.8 Variable latente et cadre épistémologique

La démarche épistémologique discutée précédemment peut être qualifiée de démarche pratique, au sens où le pragmatisme-réaliste l'entend. Cette démarche n'est donc pas un

recours au relativisme dans le sens où tout serait justifiable, car la validation scientifique se base sur la confrontation entre la théorie et l'empirie. Dewey (1925) parlait *d'assertibilité garantie*, c'est-à-dire d'un protocole de validation basée sur l'empirisme et le « consensus » d'un champ scientifique. Le « consensus » développé par le pragmatisme renvoie non « à la majorité » mais en ce « *qui ne peut être réfuté par personne [...] ce consensus n'a donc pas le sens d'un consensus de droit, mais de fait* » (Pereira, 2012). Une telle démarche pragmatique permet de contourner la tension entre « réalité » et « subjectivité » contradictoire avec le réalisme-empirique qui impose une réalité « objective » inerte à découvrir. De même elle permet d'éviter la contradiction du constructivisme en ne considérant pas les propriétés mentales que sous l'angle de la construction sociale.

Les concepts liés aux propriétés mentales permettent de caractériser des propriétés structurantes de la personne, c'est-à-dire des propriétés ayant une homogénéité et une stabilité suffisantes pour la personne pour être abordées comme des objets « réels » même si non-inertes (stabilité conséquence d'attracteurs stables du réseau en amont). Le cadre épistémologique pragmatique-réaliste proposé permet de considérer que les concepts développés en psychologie se réfèrent à des propriétés mentales en tant que réalités pratiques. Nous avons donc un point de vue réaliste qui est nécessaire avec les variables latentes (Borsboom et al., 2003). Un protocole de mesure²⁰ permettant l'usage de variables latentes semble donc opérationnel.

Les propriétés mentales, comme des propriétés émergentes, permettent une régression conceptuelle et méthodologique par rapport aux processus mentaux de bases. La conceptualisation de certaines de ces propriétés mentales renvoie à des catégories de propriétés qui sont en tension entre réalité et construction. Pereira (2009) « *Le découpage de la réalité est ambigu. Il semble à la fois relever du fait, donc de la réalité, et du langage, donc des valeurs culturelles* » (p. 42). Une démarche pragmatique permet de les considérer comme des objets réels même si non-indépendant du cadre pratique de la recherche. Les variables latentes comme estimateur de ces propriétés mentales à partir de leurs manifestations permettent de développer un cadre d'opérationnalisation de ces propriétés mentales. De ce point de vue, si nous affirmons une rupture avec l'empirisme-réaliste et posons clairement un cadre épistémologique « pragmatique-réaliste », les variables latentes semblent des outils opérationnalisables pour inférer des propriétés mentales.

²⁰ Le terme « mesure » est à prendre au sens large et non au sens d'une métrique attribuée. Nous reviendrons sur le problème de la mesure dans notre propos.

1.2.3 Le concept de Construit

Les articles académiques avec variables latentes usent le plus souvent du concept de « construit », concept qui a généré récemment une discussion sur son sens réel. Un consensus s'est dégagé pour considérer ce terme de « construit » comme inopérant... pour voir tous les auteurs continuer à l'utiliser. Cette discussion critique sur le concept de « construit » me permettra de caractériser de façon plus claire la démarche « pragmatique-réaliste » inhérente aux travaux sur des propriétés mentales.

1.2.3.1 Qu'est-ce qu'un « Construit » ?

La définition de Cronbach & Meehl (1955, p. 283) est la plus souvent citée comme initiatrice de la conceptualisation du construit qui est « *A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct* ». Sur cette base, le construit est une entité (théorique) inobservable qui entraîne des mesures observables. La littérature académique considère que le construit a un sens supérieur (*surplus meaning*) que l'opérationnalisation du construit par la mesure empirique. Cependant, la littérature académique récente pointe une confusion sur ce concept : les modèles avec variables latentes annihilent souvent la différence entre concept et référent (Lovasz & Slaney, 2013; Maraun, 2010, 2012; Maraun & Gabriel, 2013; Maraun & Halpin, 2008; Maraun & Peters, 2005; Maraun et al., 2005; Slaney, 2001, 2012; Slaney & Racine, 2011, 2013a, 2013b). Slaney & Racine (2013a, 2013b) montrent que la définition claire et précise de « construit » est rare en psychologie à partir de citations tirées de la littérature académique. Maraun & Gabriel (2013) et Slaney & Racine (2013b) font des historiques de mots introduits dans la littérature, notamment de la confusion entre *Concept*, *Réalité* et *Variable*, avec beaucoup de citations marquant cette confusion entre un *construit-concept* et un *construit-réalité*, et ceci dès l'émergence de ce concept par Cronbach et Meehl (Slaney & Racine, 2013b). La prise de conscience de cette confusion apparaît ainsi dans plusieurs articles rédigés par des auteurs centraux dans la discussion sur les variables latentes : Lissitz (2009) écrit : « *just what do we mean by the term construct ?* » ; Markus (2008a) confirme cette confusion : « *Wiley (2002) makes a distinction between two uses of the term construct. The first refers to what a test actually measures and the second refers to what the test developer intends the test to measure.* » (p. 64) ; Borsboom, Cramer, et al. (2009) discutent longuement aussi de cette confusion : « *we do not know what constructs are [...] Construct validity have the habit of*

using one word for two things, without clearly indicating when they mean what. In particular, the term “construct” is used to refer to (a) a theoretical term [...] (b) the property that we think plays a role in psychological reality and which we would like to obtain measures » (p. 1149-150).

La sémantique académique utilisée marque ainsi une confusion sur le terme « construit », entre les termes « concept », « variable latente » et « réalité ». Le construit est souvent à la fois le concept et le référent du concept. La littérature psychologique donne ainsi deux sens aux Construits, d'une part comme un concept théorique, et d'autre part comme une réalité. Maraun & Gabriel (2013) et Maraun & Peters (2005) rappellent qu'un construit ne peut pas être à la fois un concept et ce que le concept représente. Il ne peut être une représentation et représenté. L'empirisme-réaliste différencie le concept (constituant du langage) et le référent qui existe (constituant d'une réalité naturelle). Pour ces auteurs, un concept théorique permet d'exprimer notre connaissance d'un phénomène, c'est un composant du langage. Les concepts ont pour eux une fonction normative (« dominant », « agréable », « intelligent », « athlétique »). Le référent d'un concept peut être une réalité, mais c'est le langage qui fixe le concept, donc un concept n'est jamais « observable » ou « non observable » car il n'est pas constituant d'une réalité naturelle. Un concept peut signifier une entité et cette entité peut être caractérisé comme observable ou non observable. Les entités non-observables, comme les référents auxquels sont liés les états mentaux, sont de la classe des objets réels, donc différents de la classe des concepts qui les caractérisent. Si un concept peut faire référence à des objets ou événements qui peuvent être observables ou non, en tant que composant du langage, un concept n'a pas la structure d'une réalité. Il est créé par l'homme et dépend de lui. On peut découvrir des entités, mais pas le sens de ces entités. Le sens est donné par des formes linguistiques par l'homme.

1.2.3.2 Peut-on se passer du concept de construit ?

Certains auteurs proposent de ne plus utiliser ce terme de « construit ». Un « construit » tel qu'utilisé en psychologie n'a en effet pas de correspondant dans les autres domaines scientifique pour Michell (2013). Pour Maraun (2010) le construit montre la confusion entre l'investigation conceptuelle et empirique en psychologie et il faudrait retirer le terme « construit » pour éviter les confusions. Borsboom, Cramer, et al. (2009) se retrouvent sur cette nécessité de ne plus faire appel à ce concept de construit : *« Moreover, the word « construct » is so thoroughly infected with both meanings that we see no possibility to*

restrict its usage to indicate either sense (a) or sense (b). The only viable option, we think, is to dispose of talk about construct altogether and explicitly refer to “theoretical terms” for sense (a) constructs and “psychological attributes” for sense (b) constructs. In essence, this means that the theory and vocabulary of construct validity is abandoned entirely » (p. 152).

L’usage du terme « construit » semble pourtant une nécessité en psychologie au regard de son usage... même par ceux qui pointent son caractère confus. Par exemple, Slaney & Racine (2013b) discutent longuement de la confusion sur le terme construit et cherchent à clarifier la conceptualisation en psychologie. Or, Slaney (Slaney, 2001; Slaney, Storey, & Barnes, 2011) utilise le terme « construct » sans regard critique ni précaution. De plus, pour Slaney & Racine (2013b), le « construit » est dans le sens le plus commun un concept théorique particulier qui désigne une classe d’objets réels (ou potentiellement réel) mais inobservables, c’est-à-dire des concepts utilisés pour représenter des caractéristiques d’une réalité psychologique inobservable, Slaney & Racine (2013b) « *In the simplest and clearest sense, all that is meant is that psychological constructs are theoretical concepts reserved for designating classes of real (or potentially real) but currently unobservable theoretical entities (processes, etc.) about which we desire knowledge* ». Or, Slaney et al. (2011) considère le construit comme l’attribut inobservable que nous voulons mesurer, et non comme un concept théorique : « *It just so happens that latency has in this context been taken to represent a received view of psychological measurement according to which constructs are unobservable and may only be measured indirectly [...] the dimensionality of some measure (or some construct) of interest [...] the construct measured by the test* » (p. 266-273). De même Borsboom, Cramer, et al. (2009), qui appelaient à ne plus utiliser le terme construit, l’utilisent toujours et sans précautions particulières. Pour exemple Markus & Borsboom (2013b) utilisent clairement le « construit » comme la propriété que nous voulons mesurer et non comme un terme théorique : « *In CTM [causal theory of measurement], constructs refer to common causes (equivalently, attributes; Rozeboom, 1966) that underlie a set of item responses, so that people respond to items differently because they have a different construct score (Borsboom, 2008; Borsboom, Mellenbergh, & Van Heerden, 2004).* » (p. 54) ; à l’inverse, Schmittmann et al. (2013) utilisent le « construit » à la fois comme un concept théorique et à la fois comme un attribut que nous voulons mesurer (sans préciser clairement dans quel sens ce terme est utilisé).

Il semble que la littérature académique ne puisse dépasser le terme de construit. Peut-être, comme le dit Borsboom, Cramer, et al. (2009) : « *This double usage [du construit] has created an enormous amount of confusion, but has become so entrenched in both construct*

validity theory and methodological language that it has by now become too late to change it ». Mais la tentative de ces auteurs d'aller au-delà de ce terme et de n'utiliser que "terme théorique" ou "attribut psychologique" ne semble pas fonctionner car pas (ou peu) repris depuis, même par eux. Pour clarifier le sens de construit, des auteurs expliquent leur position dans des articles. Pour Markus (2008a) par exemple, le construit est ce que le test mesure (« *What the test measures directly corresponds to the notion of construct developed here* », p. 64), c'est-à-dire un objet mesurable qui varie dans la population et qui dépend de la population pour trouver son sens ; à la différence d'un concept qui a un sens indépendamment d'une mesure dans une population. Bagozzi (2011b) revient lui aussi sur la conceptualisation d'un construit et considère un construit comme un objet théorique qui représente un concept théorique mais qui a un sens théorique et empirique à partir des mesures, c'est-à-dire à une sorte d'interface entre la conceptualisation théorique et de la mesure empirique.

La discussion sur le concept de construit est symptomatique à mes yeux de la confusion sur la méthode en psychologiques.

1.2.3.3 Conclusion sur le concept de construit

Il semble que la littérature académique ne puisse dépasser le terme de construit. Cette contradiction entre « le concept » et « la réalité », cristallisée par le concept de « construit », peut trouver une réponse dans la compréhension d'un construit comme la conceptualisation d'un phénomène psychologique émergent. La psychologie a besoin d'un vocabulaire propre qui se construit dans une démarche pragmatique-réaliste, c'est-à-dire où la réalité est non-seulement conceptualisée subjectivement (comme dans tout champ scientifique), mais où les concepts sont fortement dépendant du prisme de la praxis du chercheur ; et où la propriété mentale elle-même peut être un objet mouvant (Hacking, 2008). Une propriété mentale, comme l'émotion ou l'anxiété, ne trouve pas de réalité physicochimique unique au niveau micro (via neuro-imagerie par exemple), et doit être, comme forme émergente, conceptualisée au niveau de ses manifestations, donc par le prisme déformant de la subjectivité du scientifique, et en permanence renégociable conceptuellement. Il y a donc toujours en psychologie une tension/mouvement entre le concept et la réalité que le terme construit cristallise à mes yeux. Le terme « construit » sert de pont entre une réalité non discernable autrement que par ses manifestations ; et ses manifestations qui sont subjectivées. Le construit comme terme synthétique cristallise le processus de connaissance pratique qui ne sépare pas la théorie de la pratique (le pragmatisme).

Ma conclusion sur le terme « construit » se veut donc orthogonale à la discussion académique. Dans le numéro spécial sur les construits de *News Ideas in Psychology* (numéro 13, volume 1, 2013), tous les auteurs, excepté Markus et Borsboom, proposent de dépasser le construit, avec des raisons différentes ; sachant que Borsboom co-écrit dans ce numéro un autre article appelant à déconstruire le construit (Lovasz & Slaney, 2013; Maraun & Gabriel, 2013; Michell, 2013; Schmittmann et al., 2013; Slaney & Racine, 2013a, 2013b) . Je propose une autre vision du concept de « construit ». Un construit est apparemment incontournable en psychologie car il renvoie à une conceptualisation d'un phénomène qui est une réalité tangible non-réductible, et appréhendé uniquement par ses manifestations. Le concept de Construit renvoie donc à cette tension inhérente aux propriétés psychologique entre le concept et la réalité réifiée ; il est ainsi associé généralement à deux notions différentes : le sens donné au phénomène auquel il se réfère (concept) et le phénomène lui-même réifié. Il caractérise en quelque sorte la démarche en tension entre objectivité et subjectivité inhérente à une épistémologie pragmatique. Je cite de nouveau Cometti (2010) « *Putnam veut montrer [...] que nos conceptions, pour indissociables qu'elles soient de notre langage et de notre forme de vie, n'en sont pas moins objectives, même si leur objectivité est une "objectivité-pour-nous"* » et Pereira (2012) « *dans le pragmatisme, la théorie est toujours pensée à partir de la pratique [...] il n'y a pas de sens à séparer les deux dans une démarche pragmatique [...] refus de séparer théorie et pratique* ».

Le construit permet de caractériser cette jonction théorico-empirique par l'usage d'un terme unique dans la construction de notre connaissance sur une propriété mentale. Hacking (2008) rappelle (donc en désaccord avec Michell (2013) qui considère que le terme « construit » n'est pas utilisé en sciences de la nature) que « *A la fin des années trente, les philosophes des sciences naturelles [...] avaient commencé à se servir du mot "construit" pour désigner des entités théoriques* », p. 67, et pour lui « *En psychologie expérimentale nous ne nous servons pas de catégories trouvées dans la nature, mais de construits dont la validité s'établit à travers nos pratiques* » (p. 68). Nous retrouvons cette tension entre « réalité » et « construction », et en conséquence la nécessité d'une approche pragmatique, pour les indicateurs statistiques dans l'espace public car ils sont eux aussi entre réalité et construction. Le construit en psychologie joue en quelque sorte le rôle des indicateurs statistiques en sociologie : « *La notion "d'indicateur statistique" apparaît comme une sorte de composite hybride, réunissant la mesure au sens des sciences de la nature, et le signifiant, au sens de la linguistique. L'efficacité de cet hybride résulte de sa souplesse et de sa malléabilité. Tantôt il est utilisé d'une façon réaliste, tantôt une distance est établie [...] L'indicateur apparaît*

comme une combinaison entre un énoncé verbal [...] et une mesure. » (Desrosières, 2008b) ou « La “réalité” n’est pas rejetée comme métaphysique, comme elle l’est par Pearson, mais elle n’est pas non plus assimilée à un “réel” directement mesurable comme l’est le réel des sciences de la nature. La réalité sociale est donc d’une nature intermédiaire, dont les indicateurs statistiques peuvent rendre compte, à cheval entre conventionnalisme et réalisme. » (Harari-Kermadec & Bruno, 2012).

1.2.4 Conclusion sur le cadre épistémologique

J’ai posé le cadre épistémologique qui me paraît nécessaire pour appréhender en psychologie les propriétés mentales. Après avoir synthétisé ma position, je discuterai ensuite dans cette conclusion de quelques questions philosophiques pour clarifier ma position. Mais j’informe que je reviendrai sur le pragmatisme-réaliste en tant que méthode par la suite, considérant n’avoir posé ici que les fondamentaux épistémologiques.

1.2.4.1 Les propriétés mentales et variables latentes

Le cerveau n’est pas un objet inerte et isolé, les propriétés mentales sont des propriétés émergentes à comprendre dans l’interaction sociale, donc en contradiction avec l’empirisme-réaliste qui recherche des objets stables/biologiques et extérieurs à la pratique sociale. Ces propriétés mentales sont structurantes des personnes à partir de manifestations, c’est-à-dire ayant des caractéristiques suffisamment homogènes pour une même personne pour être abordée comme des objets ayant une suffisante stabilité, même si non-inerte. L’émergence associée à une démarche pragmatique-réaliste permet de conserver le même cadre conceptuel malgré des variations du cerveau pour caractériser une propriété mentale identiquement dans le temps au sens où sa caractérisation sociale sera identique.

Les propriétés mentales ne peuvent être analysées en effet dans leur dimension pleine par la neuro-imagerie. *« Aucun physicien ou physiologue qui étudierait minutieusement le corps de Mozart, et tout particulièrement son cerveau, ne serait capable de prédire sa Symphonie en sol mineur d’une manière détaillée. » (Popper, 1984, p. 35).* Ou pour reprendre la critique de Quéré (2001) la tendance à réduire l’être à des processus causaux d’éléments de bases *« c’est rendre compte des états mentaux et de leurs contenus intentionnels en termes non mentaux et non intentionnels [...] Le sens est donc quelque chose que certaines choses ont (les mots, les phrases, etc.) et que d’autres n’ont pas (les nuages, les pierres, etc.). »*

Ces propriétés émergentes, permettant une régression conceptuelle et méthodologique par rapport aux neurosciences, ne peuvent se conceptualiser que par le prisme de l'intersubjectivité. Seule une démarche reconnue pragmatique permet de contourner l'impasse posée entre « réalité » et « concept » tel que l'empirisme-réaliste le pose car cette réalité est « mouvante » autant par le processus scientifique que par la fonction sociale des propriétés mentales conceptualisées.

C'est justement parce que les objets d'étude en psychologie ont à la fois une dimension « constante » et une dimension « mouvante », qu'ils sont à la fois une « réalité » et une « construction », et que le terme de « construit » permet de cristalliser cette tension. La discussion sur le terme de « construit » caractérise ce nœud ontologique en psychologie découlant du réalisme-empirique.

Ces propriétés mentales émergentes sont à formaliser à partir des manifestations, cadre de l'opérationnalisation des variables latentes. Les variables latentes peuvent donc être des outils de formalisation de propriétés mentales si on affirme que cette formalisation est inscrite dans une démarche pragmatique. Je vais revenir sur tous ces points dans ma discussion sur la mesure.

1.2.4.2 Physicalisme et pragmatisme ?

J'ai utilisé pour le moment la définition du physicalisme donnée par Kisler (J. Kim, 2009) qui définit le physicalisme comme « *une doctrine ontologique* » considérant que « *tous les objets sont physiques ou exclusivement composés d'objets physiques ; et toutes leurs propriétés sont physiques ou réductibles sans reste à des propriétés physiques* » (p. 10). Putnam (1984) donne une définition très différente « *Le "physicalisme" est la doctrine qui dit que toute la signification "factuelle" d'un énoncé peut être traduite sans résidu dans le langage de la physique* » (p. 226).

Pourtant Putnam s'affirme « naturaliste » et en rupture avec le dualisme cartésien. Si pour Putnam la pensée « *is not a thing; talk of our minds is talk of world-involving capabilities that we have and activities that we engage in* » (cité par El-Hani & Pihlström, 2002) ; cette pensée ne semble pas une non-matière au sens du dualisme. Putnam est donc en rupture avec le physicalisme réductionniste car le langage de la physique ne peut traduire le sens d'une conceptualisation telle que définit précédemment. Pour Putnam (1984) les propriétés mentales ne sont pas « physiques » car « *on peut [les] définir sans mentionner la physique ou la chimie du cerveau [...] dans la mesure où elle peut être réalisée par des systèmes différents,*

indépendamment de leur constitution ontologique [...] une même propriété psychologique peut appartenir à mille espèce différente, avec des constitutions chimiques et physiques tout à fait différentes [...] les propriétés psychologiques sont identiques aux propriétés fonctionnelles » (p. 92-93). Cela renvoie à une position proche de celle que j'ai défendue sur la nécessité d'appréhender les propriétés mentales par leurs manifestations. Mais il y a à mon avis une tension dans la définition de Putnam du physicalisme entre une définition ontologique et une définition gnoséologique. Il me paraît important de clarifier le rapport pragmatisme/ontologie des propriétés mentales au regard de la confusion sur cette « réalité ». Je considère comme Putnam que les propriétés mentales induisent un « sens » non traduisible dans le langage de la physique car ce sens est toujours contextualisé. Donc si le physicalisme renvoie au langage de la physique, c'est-à-dire en lui donnant un sens non « ontologique » mais « gnoséologique », je peux m'accorder avec Putnam et me démarquer du physicalisme. Mais il me semble alors que l'on discute du physicalisme réductionniste qui considère que l'on peut tout ramener à la science physique. A contrario, si on pose le physicalisme au sens ontologique du terme, comme Kisler, il me paraît possible de considérer une position ontologique physicaliste des propriétés mentales, mais un physicalisme non réductif renvoyant à l'émergence. Cette position ontologique physicaliste ne me paraît pas contradictoire avec une critique de la position gnoséologique du physicalisme, qui renvoie à une critique du physicalisme réductionniste : critique de la connaissance des propriétés mentales par la connaissance microbiologique. Donc si le sens d'une propriété mentale ne peut être traduit en langage physique, en accord avec Putnam ; les « propriétés » mentales n'en sont pas pour autant des objets non-physiques. Je conserve donc une position physicaliste non-réductionniste qui ne me semble pas contradictoire avec le pragmatisme développé par Putnam.

1.2.4.3 Emergence et pragmatisme

Je me suis appuyé sur l'émergence simple pour statuer la nature émergente des propriétés mentales. Mais le pragmatisme permet à des auteurs de légitimer l'émergence forte des propriétés mentales. El-Hani & Pihlström (2002) par exemple, s'appuyant entre autres sur les travaux de Putnam, combine l'émergence forte et le pragmatisme pour poser que la réalité n'est pas que physique. Ces auteurs reprennent de Putnam qu'il y a différents niveaux de réalité, donc différents niveaux de conceptualisation, et en conséquence que le cadre de la physique formel n'est pas forcément le niveau requis pour les propriétés mentales car notre

approche est toujours relative à nos pratiques, objectifs, notre connaissance et notre cadre linguistique. Pour ces auteurs, le pragmatisme donne un cadre de légitimation d'une vision émergentiste des propriétés mentales sans recours à l'émergence simple, car l'émergence forte apporte une fonctionnalité scientifique pour eux.

D'un point de vue gnoséologique, je peux être en accord. Mais d'un point de vue ontologique, cette position glisse pour moi vers un point de vue relativiste avec lequel je serai critique. En effet, l'émergence forte considère la causalité des propriétés mentales indépendamment des états physiques du cerveau (via la survenance). C'est à mon avis tout l'intérêt de l'approche de l'émergence simple qui concilie le physicalisme et la non-réduction. Le sens d'une propriété mentale est en effet caractérisé par l'interaction sociale. Donc en tant que tel, une structuration dans un état précis du cerveau ne peut avoir de « sens » car ce sens est une construction sociale à partir des catégories socialement admises... mais en même temps ces catégories conscientisées, ce sens, sont-elles mêmes des propriétés mentales des acteurs de cette interaction. On peut donc à mon avis très bien concilier une position émergentiste sans sombrer ni dans le réductionnisme ni dans le relativisme.

Il est à noter cependant que Putnam rejetait le concept d'émergence car ce concept sert aux sciences dures, et la pensée humaine ne doit pas être abordée comme une chose : « *is not a thing; talk of our minds is talk of world-involving capabilities that we have and activities that we engage in* » (cité par El-Hani & Pihlström, 2002). Visiblement pour Putnam l'émergence est une métaphore qui ne sert à rien, mais il me semble qu'il discute alors de l'émergence forte. Si nous discutons de l'émergence faible, la réification d'une propriété mentale ne me paraît pas contradictoire avec cet « engagement » propre au pragmatisme, car cette propriété mentale est considérée comme une propriété émergente inscrite dans une praxis sociale.

1.2.4.4 Conclusion sur le pragmatisme-réaliste

Si le pragmatisme-réaliste peut être associé pour certains auteurs à une démarche générale en science, d'autres circonscrivent le pragmatisme aux sciences sociales, car ayant des objets d'étude différents de ceux des sciences de la nature. Ma réflexion discute des propriétés mentales et de leur modélisation par des variables latentes ; je cadre la discussion sur ces objets précis sans prétendre apporter une réflexion plus élargie sur la science en générale.

Je reviendrai sur le pragmatisme-réaliste que je propose comme méthode en psychologie lorsque je discuterai des mesures de propriétés mentales. Je m'appuierai sur Putnam (Putnam, 1984, 1998, 2004, 2011), notamment sur ce qu'il a écrit dans « *Raison, vérité et histoire* »

(Putnam, 1984), car les travaux de Putnam offrent une perspective pragmatiste qui semble pertinente pour notre sujet ; « *néo-pragmatisme* » pour Pereira (2009) ou « *réalisme pragmatique* » pour Tiercelin (2002). Putnam a une pensée évolutive (Tiercelin, 2002), mon objectif n'est donc pas d'affirmer une adhésion à tout ce qu'il a écrit, mais de me servir de certains de ses développements afin de préciser ma thèse en développant un cadre pragmatiste-réaliste à même d'offrir un théorie de la connaissance efficiente en psychologie. Comme je l'indiquerai par la suite, mon développement a par ailleurs également des convergences avec Searle (1996) ou Hacking (2008).

Comme le critique Hacking (1992), avec raison je pense, Putnam ne discute pas de *raisonnement*, mais de la *raison* (pure) ; il n'illustre ses développements que par des *fables de sciences-fictions* (comme le dit Hacking, 1992) dans ses ouvrages (p.e. nous trouvons dans les exemples de Putnam des renvoie à un *planète jumelle de la terre* ou au *cerveau dans une cuve*), sans s'appuyer sur des faits concrets de la science comme Hacking, par exemple. Putnam discute donc d'une façon générale, posant un pragmatisme-réaliste comme cadre générique à toute démarche de connaissance. Or, mon objectif est bien de discuter des raisonnements scientifiques en psychologie. Comme déjà dit, je ne pose ma discussion que sur les propriétés mentales, sans vouloir prétendre l'élargir à tous les champs scientifiques. En ce sens, je peux considérer que mon développement est sur cet aspect plus proche de celui de Hacking que de celui de Putnam. Hacking (2008), en effet, différencie les sciences sociales des sciences de la nature, et je le rejoins lorsqu'il considère que les discussions ne sont pas identiques dans ces deux champs. Hacking adopte une position réaliste, et donc critique du pragmatisme-relativiste tel que porté par Rorty, et en conséquence il ne se définit pas comme pragmatiste (Hacking, 2007). Mais un certain nombre de travaux contemporains en philosophie des sciences l'affilie au « *new pragmatism* »²¹. De même Searle semble être considéré comme proche du courant pragmatique par les influences auxquelles il se réfère.

Je considère la position de Hacking (2007) sur le pragmatisme comme pertinente : « *one of the few domains in which I am a consistent pragmatist is pragmatism itself : use it when it is useful, but don't when it isn't* » (p.32). C'est bien parceque ce recours au pragmatisme-réaliste me semble opérant que je propose ce dernier, et non une affiliation à un courant indéfinissable par ailleurs sur le plan philosophique. Hacking critique le pragmatisme, mais il me semble qu'il se réfère essentiellement à Rorty, qui s'affirmait « *néo-pragmatiste* », en

²¹ Sur Hacking et sa proximité avec le pragmatisme voir par exemple : Aboulafia & Shook (2008) ; Misak (2007) ; Bem & Jong (2006).

développant une philosophie opposée au réalisme que Hacking critique : « *Rorty's version of pragmatism is yet another language-based philosophy, which regards all our life as a matter of conversation* » (p. 41). Hacking refuse ce scepticisme sans sombrer dans le réalisme-empirique, voyant bien une pratique dans la connaissance de la réalité : « *My own view, that realism is more a matter of intervention in the world than of representing it in word* », et critiquant chez Dewey cette absence de connaissance potentielle : « *The right track in Dewey is the attempt to destroy the conception of knowledge and reality as a matter of thought and of représentation* » (p. 41). Je peux donc tout autant me référer à Hacking qu'à Putnam pour la méthode développée, sans considérer que le nombre de références utilisées dans ma thèse caractérise un positionnement affirmé en faveur de Putnam et à l'encontre de Hacking ou de Searle par exemple.

J'ai posé le terme « pragmatisme-réaliste », en contradiction avec un pragmatisme relativiste comme développé Rorty, mais en contradiction également avec l'empirisme-réaliste. C'est toute la difficulté de la catégorisation car un terme réifie ; problématique que je développerai par la suite au sujet des propriétés mentales, mais qui peut s'étendre aux théories. J'aurais pu adopter le terme « *new pragmatism* », mais je perdais le sens de « réalisme » qu'il me paraît nécessaire d'affirmer ; de même le terme « constructivisme-réaliste » me paraît souffrir de l'affiliation au relativisme/septicisme du constructivisme.

Mon objectif est clairement de ne pas encapsuler dans un terme générique la méthode que je propose, mais bien de définir ce que doit être une méthode de travail en psychologie, car une méthode implique une démarche dynamique à l'encontre d'un terme rigidifié. Si je m'appuierai sur Putnam par la suite c'est que cet auteur a développé des aspects qui me paraissent très pertinents pour mon propos, mais cette référence à Putnam n'est pas à prendre comme une position critique de Hacking ou de Searle par exemple. Je considère ne pas avoir à « choisir mon camp » dans un certain nombre de débats philosophiques mais bien à devoir poser le plus explicitement possible la méthode qui me paraît efficiente pour les objets que je considère ici : les propriétés mentales. Je conserverai donc le terme « pragmatisme-réaliste » sans le considérer comme affilié à un courant clair en philosophie des sciences mais bien comme un terme désignant une méthode efficiente pour la psychologie.

2 La mesure des propriétés mentales

La mesure de propriétés mentales est au cœur d'une discussion importante dans la littérature académique, et renvoie à toute la démarche en psychométrie. « *The persistence of critical publications about measurement in psychology bears witness to a profound crisis at the heart of the discipline* » (Vautier et al., 2012, p. 810). La question de la mesure de propriétés mentales est donc centrale. Je vais dans un premier temps rappeler l'impossibilité de poser un cadre formel de la mesure en psychologie comme il l'est en physique; je reviendrai ensuite sur ce que la psychologie cherche à mesurer ; je développerai alors un cadre de mesure pour des propriétés mentales et les tensions avec la pratique ; pour finir sur une critique de l'usage de la mesure en psychologie.

2.1 Le problème de la mesure en psychologie

Il est hors de ma compétence de prétendre rapporter l'exhaustivité de la discussion relative aux problèmes liés à la mesure en psychologie ; des journaux académiques sont spécialement dédiés à cela (*Measurement ; Measurement: Interdisciplinary Research and Perspectives ; Journal of Applied Measurement ; Applied Psychological Measurement ; Educational and psychological measurement ; Journal of Methods and Measurement in the Social Sciences ;...*). Mon objectif est d'essayer d'apporter des éléments de réflexion et de réponse, qui permettront de cadrer la discussion dans le prolongement de ce que j'ai déjà discuté. Je la centrerai sur quelques points, en m'appuyant sur les cadres formels les plus usuels en psychologie : qu'est-ce que mesurer ? Les mesures conjointes (Luce & Tukey, 1964) sont-elles applicables en psychologie ? La *latent theory* est-elle une théorie de la mesure ? Le modèle de Rasch est-il efficient ?

2.1.1 Science pathologique ?

Tout un débat a été impulsé par les articles et livres de Michell (Michell, 2013, 1997a, 1997b, 1999, 2000, 2008a, 2008b, 2008c, 2012), des articles de Kyngdon (Kyngdon, 2008a, 2008b, 2008c, 2011a, 2011b, 2013) ou de Trendler (Trendler, 2009, 2013) ou Vautier et ses co-auteurs (Lacot, Afzali, & Vautier, 2015; Vautier, 2011, 2012, 2013, 2015f, 2015g; Vautier & Afzali, 2015; Vautier, Lacot, & Veldhuis, 2014; Vautier et al., 2012; Ziegler & Vautier, 2014)

qui discutent des erreurs scientifiques liées aux travaux avec les variables latentes et à la mesure de celles-ci.

Pour Michell (2008b), la psychométrie est une science qui souffre d'une « pathologie », car les mêmes erreurs se manifestent en permanence et les psychométriciens les maquillent car ne pouvant pathologiquement pas faire autrement. D'autres auteurs développent une critique virulente de la mesure telle que pratiquée en psychologie : Maraun parle d'« hérésie », Vautier parle de « fétichisme ». En effet, la psychométrie ne questionne pratiquement jamais l'hypothèse quantitative des propriétés qu'elle mesure. Le phénomène mesuré via des tests est considéré implicitement comme quantitatif, c'est-à-dire mesurable. Je donne, pour exemple du caractère implicite quantitatif d'une propriété mentale, un extrait de l'article très souvent cité dans la littérature de Borsboom et al. (2003, p. 204) : « *Now, concepts closely related to the latent variable have been discussed extensively. These concepts include the meaning of the arrows in diagrams of structural equation modeling (see, e.g., Edwards & Bagozzi, 2000; Pearl, 1999; Sobel, 1994), the status of a strongly related concept, namely the true score of classical test theory* » (p 204). Ici le concept (qui se mélange au score, avec donc la confusion fréquente entre “concept” et “réalité”), a par nature “un score” potentiel.

Pour justifier ce choix, les psychométriciens utilisent le plus souvent la définition de la mesure de Stevens (1946, p. 677) : « *measurement is the assignment of numerals to objects or events according to rule* »). Cette définition est très lâche et contourne l'exigence d'un cadre formel de la mesure pour pouvoir considérer une propriété comme mesurable. Michell montre que la littérature psychologique reprend le plus souvent (excepté Luce, Tuckey, Tversky, Krantz, Suppes et Narens) cette définition (Michell, 1997a, 1999). Pour exemple, le livre récent de Hair, Hult, Ringle, & Sarstedt (2013), c'est-à-dire des auteurs centraux sur les SEM, se positionne explicitement relativement à cette définition de la mesure : « *measurement is the process of assigning numbers to a variable based on a set of rules* » (p. 5).

Michell (2013) critique cette définition car on ne peut pas qualifier comme « quantitatives » des propriétés si elles ne vérifient pas le cadre formel d'une structure quantitative « *Attributes are only measurable if they are continuous quantities, a feature involving specific empirical conditions. Before anyone can safely conclude that hypothesised attributes possess continuous quantitative structure, it is necessary to consider evidence diagnostic of quantitative structure and presently, such evidence is lacking (Michell, 2000, 2008). While anyone is free to consider the hypothesis that some particular attribute possesses this sort of structure, there is no scientific advantage to be gained by postulating its truth à priori.* ».

Pour Michell, cette pathologie a une origine : le présupposé est que la réalité est fondamentalement quantitative (approche pythagoricienne de la réalité) et que la démarche scientifique doit se rapprocher des mathématiques (Michell, 1999). Il existe une sorte d'impératif quantitatif en science dont la psychologie est victime ; je reviendrai sur ce point ultérieurement.

2.1.2 Le cadre formel classique de la mesure

Une mesure peut être appréhendée d'une manière formelle, avec en référence le cadre formel de la physique. Dans ce cadre, Michell (1997a) rappelle que l'attribut d'un objet est mesurable s'il y a une structure quantitative exprimable en termes de ratio. Le système axiomatique associé à la théorie de la mesure peut se formaliser par les axiomes donnés par (Kyngdon, 2008b; Michell, 1997a) :

Soit A l'ensemble des objets, et R une relation entre ces objets. Une structure empirique $\mathcal{A} = \langle A, R, \circ \rangle$ où \circ représente la relation de *concaténation* des objets (combinaison physique de deux objets). Il y a un homomorphisme entre $\mathcal{A} = \langle A, R, \circ \rangle$ et la structure relationnelle entre valeurs numériques $\mathfrak{N} = \langle \mathbb{R}^+, S, + \rangle$, où S est le jeu de relations et la relation « $>$ » représente la fonction « est plus grand ». On a l'axiomatisation suivante : (cf. Kyngdon, 2008b, p. 91) ; on retrouve une même axiomatisation dans les articles de Michell, mais la fonction « $+$ » est alors utilisée au lieu de « \circ ». Le terme « \circ » permet de généraliser la forme relationnelle :

E1 : Pour tout a, b et c dans A , $>$ est transitif;

E2 : Pour tout a, b, c et d dans A , $>$, si $a > b$ et $c > d$, alors $a \circ c > b \circ d$

E3 : Pour tout a, b et c dans A , $a \circ b > a$ et $a \circ b > b$

E4 : Pour tout a, b et c dans A , $a \circ (b \circ c) \sim c \circ (a \circ b)$

E5 : Pour tout a et b dans A , si $a > b$ il existe c tel que : $(b \circ c) > a$

E6 : Pour tout a et b dans A , il existe un entier n tel que $a(n) > b$

Si un attribut a une telle structure, il est par principe mesurable. On peut attribuer un ratio r entre a et b : $a = r \times b$, a est mesuré dans l'unité de b . Dans le prolongement de ce cadre formel, Campbell définit la mesure comme étant l'attribution de nombre pour représenter les propriétés quantitatives de ces objets.

La mesure est donc la découverte des ratios entre magnitudes. De par un tel formalisme mathématique, en 1940, la *British Association for the Advancement of Science* a conclu que,

puisque les attributs psychologiques ne vérifient pas la concaténation, ils ne sont pas des attributs quantitatifs et ne peuvent être mesurés scientifiquement.

2.1.3 L'approche de Stevens

En réaction, Stevens (1946) proposa une conception de la mesure adaptée à la psychologie. Stevens reprend la définition de Campbell mais en retirant l'exigence de l'ordre et de l'additivité. Pour Stevens (1946, p. 677) : « *measurement is the assignment of numerals to objects or events according to rule* ». Les propriétés d'additivité et d'ordre étant alors attribuées à des mesures spécifiques : d'intervalles et de rapport.

Je renvoie au livre de O. Martin, « La mesure de l'esprit » (O. Martin, 1997), où plusieurs chapitres expliquent la genèse et l'usage de la mesure proposée par Stevens qui avait une vision bien plus approfondie de la mesure que ce que laisse croire sa définition critiquée par Michell. Mais le fait est que le plus fréquemment, les chercheurs usant de mesures quantitatives, se réfèrent à Stevens sans questionner si la propriété étudiée peut être mesurée (Michell dit n'avoir jamais trouvé un seul article justifiant la quantification d'une propriété mentale).

Ce n'est donc pas tant l'approche de Stevens qui paraît critiquable, que son usage sans modération, laissant croire aux psychologues qu'ils peuvent mesurer sans questionnement. Si nous reprenons les mots de Desrosières (Desrosières, 2014), il y a dans la démarche usuelle en psychologie, basée sur la définition de Stevens de la mesure, une confusion entre « mesurer » (qui est d'attribuer une valeur à une propriété réelle mesurable) et « quantifier » (qui est d'attribuer une valeur selon une « *convention d'équivalence* »). Ainsi, avec la définition de Stevens comme centrale, l'approche de la mesure en psychologie devient clairement opérationnaliste, *quantifiant* plus que *mesurant* des propriétés mentales. Pour Michell (1999) on confond alors l'action de mesurer et l'attribut qui est mesuré. La propriété mentale correspond à ce que le test mesure sans que la psychologie ne puisse justifier cette démarche.

2.1.4 La théorie représentationniste de la mesure

La théorie représentationniste (RTM) a été développée dans les années 50 pour répondre à l'impossibilité d'utiliser des mesures dans le cadre formel classique en psychologie (Luce & Suppes, 2002). La RTM prolonge la définition de mesure posée par Stevens (Leplège, 2003a;

O. Martin, 1997). La RTM donne un cadre formel pour comprendre des phénomènes naturels par une représentation en termes de structures mathématiques, le plus souvent des nombres réels : « *intuitively, such representations seems a possibility when dealing with variables for which people have a clear sense of "greater than"* » (Luce & Suppes, 2002, p. 1). Cette théorie renvoie à une littérature abondante, dont notamment les trois tomes de *Foundation of measurement*, assez techniques (Krantz, Luce, Suppes, & Tversky, 2006a, 2006b; Suppes, Krantz, Luce, & Tversky, 2006). Luce & Suppes (2002) ont par ailleurs fait un chapitre de synthèse sur cette théorie de la mesure. Je m'appuierai sur ces ouvrages, mais sans développer les aspects techniques de cette approche, exceptés pour ceux qui renvoient à l'analyse conjointe, sujet de la section suivante, car l'analyse conjointe sert de cadre formel de référence pour la psychologie, entre autre pour le modèle de Rasch dont je discuterai ensuite.

La RTM donne un cadre axiomatique pour comprendre des phénomènes naturels par une représentation en termes de structures mathématiques, le plus souvent des nombres réels. Mesurer sur une échelle quantitative implique d'avoir un homomorphisme entre la propriété étudiée et les valeurs attribuées (et pas forcément un isomorphisme, car deux entités différentes peuvent exhiber un même état de la propriété, et donc renvoyer à la même valeur numérique ; mais par la suite comme Luce et Suppes j'utiliserai l'isomorphisme comme référent, car plus clair). Il faut, d'une part, vérifier « l'existence » d'un tel isomorphisme : le passage d'une structure de la propriété à une échelle numérique permet-il d'avoir une représentation numérique fidèle à la structure de la propriété ? Il faut aussi vérifier l'unicité de cette représentation, c'est-à-dire concrètement qu'il n'y a pas différentes valeurs possibles pour une même procédure de mesure.

La théorie représentationniste se différencie du cadre classique car elle considère que la concaténation ne doit pas obligatoirement être démontrée pour qu'une propriété soit représentée par une mesure, notamment grâce aux mesures conjointes développées ci-après. Plus fondamentalement, la différence entre le cadre formel classique et la RTM est la réalité de la mesure vs la représentation d'une réalité par une mesure. Michell (2008a) conteste la nature abstraite des nombres tel que posée par la RTM. Il donne cette définition : « *la mesure est l'estimation ou la découverte du rapport entre une grandeur d'un attribut quantitatif et une unité de cette grandeur* » (Michell, 1999), où donc la mesure n'est pas une construction, mais une réalité inférée. Pour la RTM, une telle structure mathématique peut représenter la structure de l'attribut naturel mais ne correspond pas à une réalité au sens du cadre formel de la découverte de rapport entre des propriétés ; les nombres sont des abstractions, et non des objets empiriques. Kyngdon (2008a) ou Körner (1962) discutent que des nombres, comme les

nombres complexes, sont des objets nécessaires sans correspondances « réelles » (au sens de la réalité, et non du corps des réels bien évidemment).

Le courant dominant se réfère aujourd'hui à la théorie représentationniste, car les mathématiques sont considérées comme un processus de représentation nécessaire à la pratique scientifique. Ainsi, sans être constructiviste, les valeurs numériques sont considérées comme représentant des propriétés réelles d'un objet dans un symbolisme abstrait que sont les mathématiques. La RTM présente le cadre axiomatique nécessaire pour considérer une telle représentation de la réalité comme fiable : « *A homomorphism into the real numbers is often referred to as a scale in the psychological measurement literature. From this standpoint measurement may be regarded as the construction of homomorphisms (scales) from empirical relational structures of interest into numerical relational structures that are useful.* » (Krantz et al., 2006a).

2.1.5 Les Mesures Conjointes

La RTM renvoie le plus souvent aux mesures conjointes. Ces dernières vont apporter une « révolution » par rapport au paradigme dominant sur la mesure.

On différencie généralement les attributs dont la qualité est une mesure extensive (où la concaténation est directement mesurable) comme la longueur d'un bout de bois ; et les attributs dont la qualité est une mesure intensive (où la concaténation est non directement mesurable) comme la densité ou la température. Les mesures intensives découlent de mesures sur d'autres attributs que l'attribut mesuré : la densité découle du volume et de la masse par exemple. Elles se réfèrent à des attributs où la concaténation entre deux éléments ne se fait pas de façon « additive », au sens du signe mathématique « + » (concaténer deux volumes identiques d'eau ayant des températures différentes ne donne pas un volume d'eau ayant une température égale à la somme des deux températures). Il faudrait donc, selon le formalisme usuel de la mesure, valider le caractère quantitatif des attributs élémentaires pour pouvoir caractériser le caractère quantitatif de l'attribut dérivé. Les propriétés mentales sont considérées comme des propriétés « intensives » ; de plus on ne peut prouver le caractère mesurable des éléments qui la composent (les processus en amont). La théorie des mesures conjointes va apporter une révolution sur ce point : permettre de considérer une propriété intensive comme mesurable (vérifiant les axiomes nécessaires du cadre formel de la mesure), sans devoir prouver auparavant que les propriétés de bases sont elles-mêmes mesurables.

Attribuée à Luce & Tukey (1964)²², cette théorie apporte un cadre formel pour valider le caractère quantitatif d'attributs dont la mesure est intensive. Cette théorie discute de situations où deux attributs A et X, non-forcément connus comme quantitatifs mais respectant une relation d'ordre (bi-ordre), sont liés à un troisième P (lui aussi non-forcément connu comme quantitatif). Les mesures conjointes permettent, par des relations spécifiques vérifiables empiriquement, de caractériser les trois attributs comme quantitatifs, sans devoir recourir au formalisme de Hölder. Luce & Tukey (1964) démontrent ainsi que des attributs non-capables de concaténation peuvent être quantitatifs, sans devoir vérifier le caractère quantitatif des attributs de base. Cette théorie permet de vérifier ainsi la qualité « quantitative » d'attributs de façon indirecte, et en conséquence, permet de considérer des attributs, considérés comme non-concaténables comme potentiellement quantitatifs. Les mesures conjointes additives permettent ainsi, à partir d'une connaissance faible des données (simple relation ordinale), de prouver le caractère quantitatif des attributs en question.

2.1.5.1 Le cadre formel des mesures conjointes

L'objectif des mesures conjointes est donc de mesurer des attributs non-extensifs (pas de concaténation empirique possible), tout en se référant à la théorie formelle de la mesure. On recherche une relation d'additivité pour un attribut par la relation conjointe de deux autres attributs. Une structure conjointe consiste en une structure d'ordre qui peut être décomposée en au moins deux sous-structures ordonnées : soit deux attributs A et X, formés de a et de x . $A \times X$ constitue toutes les paires (a, x) qui représentent l'attribut P pour lequel on veut vérifier une structure d'additivité. Pour que $A \times X$ vérifie les caractéristiques des mesures conjointes additives il faut : indépendance entre A et X ; la *simple cancellation*²³ et la *double cancellation* pour caractériser une relation d'ordre. La nature additive se montrera par deux autres axiomes.

2.1.5.2 Simple cancellation

A et X sont des attributs « naturels ». a , b , et c sont trois niveaux de A et x , y et z trois niveaux de X. La relation dans P vérifie la simple cancellation *ssi* :

²² D'autres formalisations concomitantes ont été proposées à la même période, dont celle de Debreu (1959) à laquelle Luce et Tuckey font explicitement référence. A savoir que toutes les références datent ce travail de Debreu à 1960... mais sa publication première semble être 1959.

²³ Nous utiliserons le terme anglais, qui est plus compréhensible que « simple annulation ».

$$\forall a, b, d \in A \text{ et } x, y, w \in X \in, \text{ on a :}$$

$$(a, x) > (b, x) \implies (a, w) > (b, w)$$

$$\text{et } (a, x) > (a, y) \implies (d, x) > (d, y)$$

La relation d'ordre des deux niveaux d'un attribut (A) est indépendante du niveau d'un second attribut (X), et réciproquement. La relation d'ordre est transitive :

$$\text{si } (a, x) > (b, x) \text{ et } (b, x) > (b, y) \implies (a, x) > (b, y)$$

2.1.5.3 Double cancellation

La simple cancellation ne permet pas de définir une relation d'ordre entre tous les éléments de P. La simple cancellation est nécessaire mais non-suffisante pour considérer A et X comme quantitatifs :

si $(a, x) > (b, y)$ on ne sait pas la relation entre (a, y) et (b, x) .

La double cancellation est un axiome sur les relations au sein de P :

$$(a, y) > (b, x) \text{ et } (b, z) > (c, y) \implies (a, z) > (c, x)$$

Cette relation n'est vraie que si :

$$a + y > b + x \text{ et } b + z > c + y$$

Cette relation n'est donc vraie que si A et X vérifient l'additivité, donc sont quantitatifs. Si la double cancellation n'est pas vraie, on ne peut conclure sur le caractère quantitatif de A et X.

Tableau 2 : Schéma de la double cancellation

	x	y	z
a	a,x	a,y	a,z
b	b,x	b,y	b,z
c	c,x	c,y	c,z

Il faut vérifier cette relation pour tous les éléments de la diagonale du tableau précédent, si A a n éléments et X m éléments, le nombre d'instances de double cancellation est $n! \times m!$

Si la simple cancellation a été testée, le nombre d'instances réellement à tester est $\binom{n}{3} \times \binom{m}{3}$ où n et m sont le nombre d'éléments de A et X.

2.1.5.4 La solvabilité et l'axiome archimédien

La simple et double cancellation ne définissent pas une relation quantitative pour P. Il faut que les éléments répondent à deux autres axiomes :

- La solvabilité est définie de la sorte. Il faut montrer que : $\forall a, b \in A$ et $\forall x \in X, \exists y \in X$ tq : $ax = by$
- La condition archimédienne : Soit I un ensemble d'entiers consécutifs, les niveaux de A forment une séquence standard ssi il existe x et y dans X ($x \neq y$) tels que pour tout i et $i+1$ dans I on a : $(a_i, x) = (a_{i+1}, y)$

Ces deux derniers axiomes posent en quelque sorte une *forme de densité* (Luce & Suppes, 2002). Par les mesures conjointes, on permet ainsi d'associer une représentation des objets non-concaténables par des nombres réels. Le problème est que ces deux derniers axiomes demandent une infinité de résolutions, ce qui est donc empiriquement impossible à vérifier. Dans les cas simples avec trois niveaux pour A et X, la double cancellation suffit ; avec quatre niveaux il faut ce qu'on appelle la « triple cancellation » (je ne développe pas et renvoie à Scott, 1964). Mais concrètement, dans les situations générales, on ne peut pas aisément démontrer que nos données vérifient ces axiomes.

2.1.5.5 L'erreur de mesure

Le problème avec les mesures conjointes est la gestion des erreurs de mesures. Le modèle usuel ne permet pas d'intégrer l'erreur de mesure, ce qui fragilise leur utilisation en psychologie car, comme le rappelle Sijtsma (2012a, 2012b), il y a toujours en psychologie des erreurs de mesure, mais aussi des erreurs systématiques propres aux protocoles. De ce fait, les mesures empiriques, mêmes celles d'attributs quantitatifs, peuvent en psychologie violer les axiomes d'additivité (comme la transitivité par exemple) à cause de ces erreurs. Cela rend la RTM incompatible avec des mesures empiriques (Scholten, 2011b; Sijtsma, 2012b). Luce (1977, 1997) avait déjà averti que les choses se compliquent avec les mesures conjointes si les données sont bruitées, et Luce écrivait que l'on « *ne sait pas faire* ». Depuis, des modèles probabilistes semblent avoir été proposés pour les mesures conjointes (Luce & Suppes, 2002), mais ne semblent pas utilisés et restent non-discutés dans la littérature, de fait toute la littérature récente sur l'analyse conjointe les ignore.

2.1.5.6 Conclusion sur les mesures conjointes

Les mesures conjointes donnent un cadre formel de validation de la nature quantitative d'une propriété sans devoir recourir au formalisme de Hölder. Ainsi, différents auteurs, notamment Michell, appellent les praticiens à utiliser les mesures conjointes pour valider leurs mesures. Mais le bruit causé par les erreurs de mesure rend cette vérification délicate. En conséquence, sauf à de rares exceptions, les mesures conjointes ne peuvent apporter la validation formelle de la structure additive d'une propriété mentale.

Indépendamment du problème des erreurs de mesures, du fait de la difficulté de vérifier les axiomes de solvabilité, les conditions archimédiennes, ainsi que la simple et double cancellation, très peu de travaux utilisent les mesures conjoints. Cliff (1993) remarquait il y a 20 ans que les mesures conjointes n'étaient jamais utilisées... et cela ne semble guère avoir changé depuis pour les raisons évoquées.

2.1.6 Latent variable theory

Pour répondre à l'incapacité d'utiliser le cadre formel de la mesure ou les mesures conjointes pour valider les mesures effectuées en psychologie, Markus & Borsboom (2013a) proposent un cadre théorique de la mesure basé sur les variables latentes, *latent variable theory*. Je développe une critique de cette approche, car elle est en réalité plus ou moins formellement, celle qui est appliquée usuellement en psychométrie, Markus et Borsboom ne donnant qu'un cadre formalisé de cette pratique usuelle. Ce que Markus & Borsboom (2013a) proposent est assimilable à l'approche dite « *statistique* » de la mesure, voir Borsboom & Mellenbergh (2004); Borsboom & Scholten (2008); Sijtsma (2012b). Cette approche se veut différente, pour Markus et Borsboom, de l'approche formelle (Campbell), axiomatique (Krantz et al.) ou instrumentaliste (Stevens). Cette approche s'appuie sur la fiabilité statistique des modèles posés pour considérer que la mesure est efficiente. Je reprendrai dans cette partie le terme « attribut » plutôt que « propriété mentale » car c'est le terme employé par ces auteurs, et qui correspond à leur vision empiriste-réaliste (la propriété mentale est un attribut au sens d'une entité biologique).

2.1.6.1 Le cadre du modèle de mesure

Le cadre des modèles avec variables latentes pose qu'un attribut a une certaine structure et que des variables observées dépendent de cet attribut. Si plusieurs mesures dépendent d'une

seule variable latente et qu'une relation fonctionnelle est supposée, le modèle peut être estimé et testé (ce qui le rend la falsifiabilité possible pour Markus et Borsboom). L'hypothèse de l'indépendance locale est impérative... et non testable en soi, mais considérée comme réelle. Je reviendrai sur tous ces points et les expliciterai lorsque je discuterai d'une procédure de mesure ; pour le moment je ne pose que le cadre tel que Markus et Borsboom l'ont posé dans leur livre.

Le modèle de mesure est le modèle avec variables latentes qui pose une relation probabiliste entre l'attribut et ses manifestations :

$$E(X|\theta) = \alpha\theta + \epsilon \quad (3)$$

La raison d'utiliser des modèles probabilistes avec les variables latentes vient du fait qu'il n'y a en général que très rarement des modèles déterministes en psychologie. La nécessité de l'introduction d'une probabilité dans les modélisations en psychologie est liée aux bruits inhérents à tous les protocoles en psychologie pour Markus et Borsboom (voir Sijtsma (2011) pour les erreurs de mesures aléatoires et systématiques en psychologie), et c'est la grande force des modèles avec variables latentes que de poser ces erreurs dans leur formalisation. C'est là la raison du succès des modèles avec variables latentes, qui permettent un cadre de formalisation d'une mesure avec l'intégration d'un bruit, donc différent des modèles déterministes en physique. Mais comme le soulignent Markus et Borsboom, en pratique, la physique intègre aussi des erreurs de mesures (ne serait-ce qu'en conséquence de la précision des instruments de mesure). Le modèle formel n'est qu'une idéalisation, et Markus et Borsboom (p. 75) illustrent cela par la mesure d'une balance, où un modèle avec du bruit peut être utilisé, c'est-à-dire un modèle probabiliste de la mesure. Notons que bien d'autres auteurs renvoient à ce rapport entre pratique et modèle idéal pour la mesure (Körner, 1962; Sherry, 2011; Sijtsma, 2012b), mais je reviendrai aussi sur ce point lors du développement d'un cadre formalisé de mesure de propriétés mentales.

L'intégration d'erreurs de mesure divise, pour Markus et Borsboom, les représentants de la théorie représentationniste, qui considèrent d'un côté que des modèles probabilistes interdisent les mesures ; et de l'autre, les tenants de la *latent variable theory* considèrent que des modèles probabilistes ne sont pas contradictoires avec un cadre formel de la mesure. Pour Markus et Borsboom, le cadre classique est « neutre » sur ce point. A leurs yeux, l'introduction d'erreurs de mesure et leur formalisation dans le cadre d'un modèle avec variables latentes n'est donc pas forcément contradictoire avec le cadre axiomatique de la mesure : « *The difference between deterministic axiom systems, as typically presented in*

representational measurement theory, and probabilistic models, as typically presented in psychometrics, may be less principled than it seems at first sight. In particular, while representationalism does take axiomatic structures to be error-free, this does not imply that the axioms cannot apply to probabilities. As long as these probabilities are themselves in accordance with the relevant axiomatic system, they can be used to construct a scale in essentially the same way as is done in representationalism » (p. 75).

Pour Markus et Borsboom, il ne faut pas avoir une position “*essentialist*” de la mesure (p. 68). La théorie classique impose qu’il y ait une unité ainsi qu’une procédure empirique de mesure dans cette unité, ce qui est impossible à déterminer en psychologie. De plus, le terme « mesure » est réservé à priori aux propriétés quantitatives dans la théorie classique, alors qu’en psychologie, ou dans d’autres sciences, le terme « mesure » s’applique également à des propriétés nominales et ordinales (comme Stevens l’a proposé). Markus et Borsboom considèrent qu’il est plus efficient d’élargir la notion de mesure, comme le fait la théorie représentationniste. Cependant pour éviter toute confusion ils proposent de conserver le terme « *measurement* » pour les propriétés quantitatives, et le terme « *assessment* » (« évaluation ») pour le cas général.

Markus & Borsboom (2013a) considèrent leur approche comme « intermédiaire » ou « hybride » entre le cadre formel et le cadre de la RTM, car leur approche reprend de la RTM la généralisation de la mesure, mais leur cadre reprend du cadre formel la possibilité d’introduire des erreurs de mesure. La validation est basée sur une procédure empirique (la validation du modèle statistique). Pour Markus et Borsboom, « *The classical theory of measurement deals with the definition and detection of quantitative attribute structure. Representationalism handles the definition and representation of a more general class of attribute structures. Latent variable theory deals with the analysis of data that imperfectly reflect such structures* » (p. 79), considérant que : « *measurement is an ideal to which many test aspire but one that few attain* » (p. 16).

2.1.6.2 La validation d’un modèle de mesure

Markus et Borsboom écrivent : « *when a number of observed variables depend on one and the same latent variable, and a functional relation with the latent structure is assumed for each of them, then the parameters of the model can be uniquely estimated; in this case, the model also has testable consequences, which render it falsifiable* ». Le terme de falsifiabilité renvoie en général à la falsifiabilité poppérienne, mais ces auteurs ne précisent pas ce qu’ils entendent

par « falsifiabilité » car, comme je le développerai ultérieurement, la falsifiabilité peut être comprise de façon différente selon les auteurs.

Or, par ailleurs, Markus et Borsboom rappellent que “mesurer”, ce n’est pas avoir un modèle valide statistiquement ; ils rappellent que l’objectif prédictif d’un test ne peut être la seule base des résultats statistiques, car on serait alors instrumentaliste sans discuter de la nature réelle de l’attribut. La base de la mesure doit aussi être la relation causale, et le « *meaning* » attaché à l’attribut. Mais pour eux, « *the evaluation of measurement model is scientifically accessible by statistical means* » (p. 40), renvoyant au numéro spécial de *Journal of Mathematical Psychology* (2006, issue 2), qui développe des outils permettant de sélectionner des « meilleurs modèles » au sens statistique, basé sur les indices statistiques de validation ; l’éditorial introduisant ainsi ce numéro spécial : « *it is of vital importance to quantitatively compare different models for the same set of data [...] model selection is concerned with the delicate tradeoff between simplicity and goodness-of-fit* » (Wagenmakers & Waldorp, 2006). En réalité, les articles auxquels renvoient Markus et Borsboom ne discutent que de la « sélection » de modèles (via des indicateurs de GOF) sans poser de réelles cadres de falsifiabilité.

Bamber & Van Santen (2000) ont proposé dans cette même revue un cadre de validation formelle d’un modèle statistique : « *A parametric model is said to be identifiable if, given any experimental outcome in accordance with the model's prediction, there is only one no-parameter model, created from the parametric model that is consistent with the experiment's outcome. If there can be multiple such no-parameter models, the parametric model is said to be unidentifiable* » (Bamber & Van Santen, 2000, p. 21), ce qui n’est pas le cas des modèles avec variables latentes. Les variables latentes sont des modèles non-identifiables. Par ailleurs, un modèle est dit « quantitativement testable » si la validité empirique du modèle a une probabilité quasi-nulle d’être effectuée par hasard, ce qui n’est pas le cas avec des variables latentes. En effet, la littérature montre que des données simulées ne respectant pas un modèle avec variables latentes peuvent valider les critères d’acceptabilité du modèle. Un modèle non-identifiable peut être testable même s’il ne l’est pas « quantitativement » (Bamber & Van Santen, 2000). Un modèle est dit “qualitativement testable” s’il est testable, c’est-à-dire si un jeu de données peut montrer son inconsistance. On touche ici le problème de l’erreur de mesure pointée entre autres par Vautier (2015f) ou Vautier et al. (2014). L’erreur de mesure intégrée aux modèles permet d’accepter tout modèle pour peu que l’on accepte des erreurs importantes. Dans la pratique du modèle de Rasch, par exemple (cf. la discussion qui suit sur le modèle de Rasch), des situations particulières ne respectant pas un modèle à priori ne

permette pas sa falsifiabilité puisque l'erreur de mesure sert à justifier cet écart ; c'est même le plus fréquent que des données particulières ne respectent pas exactement un modèle. Les variables latentes ne sont pas « quantitativement testables », ni « qualitativement testables » selon le cadre axiomatique proposé par Bamber et Van Sauten. Cela ne veut pas dire que des données anormales ne renverront pas des interrogations sur le modèle, mais la pratique des modèles avec variables latentes n'est pas construite sur la falsifiabilité au sens où un modèle est rejeté dès qu'un jeu de données le contredit. En effet c'est un rapport « théorie-empirisme » qui pilote, et non un cadre formel de falsifiabilité. Je reviendrai sur ce point dans la partie suivante.

Il y a donc à mon avis chez Markus et Borsboom une contradiction entre un positionnement clairement critique du supra de la statistique comme procédure de validation, et dans même temps, une volonté de poser un cadre formel de la mesure basé sur la statistique. Car fondamentalement, la mesure n'est justifiée pour ces auteurs qu'au travers de la validation du modèle statistique, et nullement en interrogeant la propriété mentale en amont du modèle. Or, c'est au niveau de la propriété mentale qu'il faut discuter et valider sa mesure potentielle. A aucun moment ces auteurs ne discutent d'une procédure permettant de justifier la représentation d'une propriété mentale sur un continuum numérique avant de modéliser la mesure dans un modèle avec variables latentes.

2.1.6.3 Conclusion

En conclusion, la volonté de donner un cadre formel de la mesure, basée sur une théorie testable, me paraît très problématique tel qu'il est posé par Markus et Borsboom. Leur justification in fine (p. 78) se repose sur l'intérêt pratique des scores. Pour eux, « *Does this invalidate the use of test score? The answer to this must be: no, as long as no measurement interpretation is made* ». Mais justement, que veut dire « ne pas interpréter le résultat d'un score comme une mesure » ? Pour eux un score sert aux « *prediction, generalization, and selection* ». De telles pratiques se basent bien sur une interprétation du score ; on revient donc en quelque sorte à la justification classique : « cela fonctionne », sans poser les soubassements théoriques. D'ailleurs, dans un autre article, de ces deux mêmes auteurs, c'est exactement leur argument (voir Markus & Borsboom, 2012). Et en contradiction, ils essaient de donner des bases formelles de validation.

A mon avis nous sommes toujours en tension dans leur approche : soit il y a réellement un cadre de mesure formel, permettant de définir si on peut mesurer une propriété mentale et

comment le faire ; soit un tel cadre n'existe pas, et affirmons alors que la psychologie rompt avec le formalisme scientifique dérivé de la physique. Mais sans rompre avec le cadre formel de l'empirisme-réaliste, on en revient à Stevens et à une pratique purement instrumentaliste de la mesure, donc sans fondement « scientifique » dans ce cadre épistémologique.

2.1.7 Le modèle de Rasch

Je développe succinctement un certain nombre de discussions autour du modèle de Rasch qui est parfois considéré comme le seul modèle en psychologie vérifiant le cadre formel de la mesure (Leplège, 2003b). Cette partie fait référence à des points que je développerai ultérieurement : les modèles statistiques et les probabilités.

Rasch propose un modèle à deux paramètres : *la difficulté* et *l'habilité* pour chaque item. La difficulté et l'habilité sont indépendantes dans son modèle. La probabilité pour un individu d'habilité s de répondre à l'item de difficulté i est :

$$P(s, i) = \frac{e^{s-i}}{1 + e^{s-i}}$$

La grande différence du modèle de Rasch avec bien des modèles d'inférence, est que l'on ne cherche pas à adapter un modèle aux données, on sélectionne les données qui correspondent au modèle théorique (Leplège, 2003a, 2003b). De plus « *Une caractéristique essentielle du modèle de Rasch est qu'il a été dès l'origine conçu dans la perspective de la mesure des individus et non de celle des groupes de sujets* » (Leplège, 2001, p. 92).

Le modèle de Rasch correspondrait aux conditions des mesures conjointes car la variable dépendante (la probabilité de la réponse) est une variable mesurable basée sur deux variables indépendantes (habilité et difficulté) vérifiant l'indépendance, la double cancellation, la solvabilité, l'axiome archimédien (voir Kyngdon, 2008b, pp. 97–98) ou Scholten, 2011a, pour un rappel des démonstrations dans la littérature académique). Le modèle de Rasch produirait une relation d'ordre au minimum, et, dans le cas usuel, une unité et l'additivité ; ce qui fait qu'il produirait des mesures répondant aux exigences des mesures conjointes. Enfin, basé sur un cadre probabiliste, le modèle de Rasch vérifie *l'invariance* (« *La comparaison entre deux individus doit être indépendante des individus particuliers qui ont servi à la comparaison [...] symétriquement, toute comparaison entre deux individus doit être indépendante des stimuli particuliers qui ont été utilisés pour la comparaison* », Rasch cité par Leplège, 2003b, p. 549).

2.1.7.1 Modèle de Rasch et probabilités

Si formellement, le modèle de Rasch vérifie le cadre des mesures conjointes (Kyngdon, 2008b; Newby, Conner, Grant, & Bunderson, 2009; Scholten, 2011b), il y a un problème soulevé par Kyngdon (2008b). Kingdon écrit : « *As the 'latent trait' θ is assumed to be a subset of the real numbers* » (p. 99), s'appuyant sur Junker & Sijtsma (2001) et Karabatsos & Sheu (2004). Ceci entraîne pour Kyngdon que les structures de la difficulté et de l'habilité sont aussi des nombres réels et ne contiennent pas d'objets empiriques (en référence à la théorie de la représentativité sous-jacente aux ACM).

Le modèle de Rasch décompose une structure de probabilités (qui sont des entités mathématiques) à partir de paramètres individuels et des items (aussi des entités mathématiques) et met ainsi en relation uniquement des systèmes numériques au lieu de lier une relation empirique à un système réel, perceptible par une valeur numérique. De ce fait, pour Kyngdon, le modèle de Rasch n'est pas assimilable à une mesure conjointe car ce dernier se réfère, dans sa définition, à des objets *perceptibles* (« *noticeable* »), et ne se réfère pas à des objets abstraits comme des valeurs numériques. Or les probabilités ne sont pas des propriétés perceptibles pour Kyngdon (2008b). Si une probabilité ne peut être considérée comme « réelle », nous sortons, selon lui, du cadre des mesures conjointes car les mesures conjointes permettent uniquement de caractériser comme mesurable un attribut réel d'un objet, et non une variable abstraite conjointe.

Peu d'auteurs semblent avoir apporté une réponse à Kyngdon en dehors de Borsboom & Scholten (2008) et Sijtsma (2012b). Ce qui est intéressant à mon sens est la nature des arguments de Borsboom & Scholten (2008) et de Scholten (2011b). Je reviendrai sur la nature des probabilités ultérieurement, je ne fais ici que poser les arguments de ces auteurs. Pour Borsboom & Scholten (2008), une probabilité n'est pas forcément une pure abstraction numérique. Pour eux, le fait qu'une probabilité ne soit pas une entité observable n'entraîne pas qu'elle n'existe pas ; une interprétation des probabilités peut les considérer comme des entités réelles. Certes une probabilité ne peut se « toucher » et ne peut être qu'estimée, mais cela peut permettre de décrire une structure réelle (« *we cannot stub our toe on a probability; we need to estimate it. But of course that is why we use statistical models. We assume the model describes a structure actually present in the world, and the statistical machinery gives us parameter estimates that are supposed to capture this structure, or at least approximate it.* », p. 114). Ces auteurs se réfèrent à une vision représentationniste faible (« *weak representationalist* »), c'est-à-dire une vision où la notion de réalité perceptible peut

représenter une réalité « en principe » perceptible, et permet ainsi de considérer que la probabilité est un objet perceptible. Si l'habilité et la difficulté sont « réelles », une relation empirique entre elles est de même réelle pour ces auteurs. Pour Borsboom & Scholten (2008) le modèle mathématique cherche à estimer des probabilités et n'est donc pas interprétable comme une entité abstraite. Kyngdon (2008a) répond qu'il faut différencier une « propension » d'un objet et une « probabilité ». Je reviendrai plus longuement sur ce point entre probabilité et propension dans une partie ultérieure, et je considère que Kyngdon a raison dans cette discussion. Les arguments de Borsboom et Scholten me paraissent très fragiles pour justifier le modèle de Rasch car eux-mêmes reconnaissent que cette approche des probabilités n'est pas dans l'esprit des mesures conjointes et que cette position entraîne des difficultés.

Si je considère que Kyngdon a raison sur la nature non réelle des probabilités, à mon sens Sijtsma (2012b) apporte à Kyngdon une réponse bien plus pertinente que Borsboom & Scholten (2008). Sijtsma rappelle qu'un modèle statistique ne correspond jamais à un modèle théorique (ce en quoi il se différencie de Borsboom & Scholten, 2008) ; je reviendrai aussi sur les modèles statistiques dans une section à suivre. Le modèle statistique de Rasch ne peut correspondre à la réalité que s'il est proche du modèle théorique, ce qui doit être discuté théoriquement. Cela revient à considérer que les probabilités du modèle de Rasch ne sont pas une « réalité », mais une mesure empirique de qualités d'un attribut latent. Pour Sijtsma (2012b) le modèle de Rasch correspond au cadre formel de l'ACM si on différencie bien le modèle théorique du modèle empirique. Cela permet à Sijtsma (2012b) de considérer, comme bien d'autres (voir Newby et al., 2009) que si on peut faire correspondre le modèle de Rasch au cadre formel de l'ACM, cela n'implique absolument pas que des données validant un modèle de Rasch prouvent la nature quantitative de la propriété mentale questionnée.

2.1.7.2 Le paradoxe Guttman-Rasch

Guttman (1944) a posé un modèle purement ordinal en fonction de l'habilité des personnes. Le modèle de Guttman est strict et basé sur un simple bi-ordre (habilité et difficulté). Or la réponse aux tests génère des erreurs (fatigue, environnement, ...) et le modèle de Rasch est un modèle probabiliste qui introduit un terme d'erreur (on estime des probabilités et non un stricte bi-ordre). Pour une revue exhaustive des modèles de Guttman et de Rasch, je renvoie à Bertrand, Ahmadi, & Heuchenne (2009). Les modélisateurs du modèle de Rasch cherchent à obtenir des données conformes au modèle théorique. Cette procédure de sélection pour

valider l'adéquation des données au modèle pose un paradoxe. Nous reprenons l'écriture de Michell (2008a) qui centre l'écriture sur le rôle de l'erreur de la mesure.

Soit X l'habilité d'une personne et Y la difficulté d'un item.

Le modèle de Guttman pose une simple relation ordinale : la réponse de la personne est correcte *ssi* $X > Y$; le modèle de Rasch pose une relation probabiliste : la réponse est correcte *ssi* $X + e > Y$.

C'est cette erreur qui différencie le modèle de Guttman de celui de Rasch pour Michell. C'est donc cette erreur qui transforme le modèle ordinal de Guttman en un modèle quantitatif de Rasch. Si l'erreur devenait nulle, le modèle de Rasch deviendrait le modèle de Guttman. La recherche d'une erreur de mesure nulle entraînerait la disparition de la qualité additive du modèle de Rasch. Michell (2008a) parle de paradoxe car la recherche de la précision fait perdre la qualité additive du test, ce qui pour Michell montre que le modèle de Rasch n'est pas un modèle répondant aux axiomes de la mesure, car seule l'introduction d'une erreur permet de considérer que le modèle de Rasch donne une mesure quantitative.

Scholten (2011a) apporte une réponse à Michell. Scholten valide le fait que le modèle de Guttman est le modèle limite de Rasch lorsque l'erreur décroît. Mais Scholten prend le problème par une autre approche. Elle discute que parfois l'erreur est bénéfique, améliorant la fiabilité de la mesure (se référant à des travaux en astronomie) et illustre par un exemple. Soit 4 personnes de 1.60 m, 1.65 m, 1.90 m et 1.95 m, si on les ordonne un certain nombre de fois (sans les mesurer) avec une bonne précision, nous aurons une statistique d'ordre quasi parfaite mais sans pouvoir inférer la différence entre les personnes. Par contre si « *un vieux professeur ayant oublié ses lunettes* » réitère le classement un grand nombre de fois, il se trompera souvent entre 1.60 et 1.65 ou entre 1.90 et 1.95, par contre il ne se trompera que très rarement entre 1.60 et 1.95. La statistique d'ordre après différentes répliques verra des objets classés fréquemment alternativement plus grand ou plus petit (1.60 et 1.65) et d'autres conservant toujours le même ordre (1.60 et 1.95), cette statistique permettra d'inférer une valeur aux différentes personnes plus aisément que lors de la première procédure. Via des simulations, elle montre qu'une certaine erreur de mesure peut être ainsi bénéfique. Cette erreur ne joue un rôle « bénéfique » que lorsqu'elle est liée à l'attribut latent (ici la taille des personnes) et n'est pas indépendante de cet attribut. Cette réponse nous paraît donc problématique car une erreur dépendante de la valeur de l'attribut latent rompt avec le cadre usuel des modèles avec variables latentes. En effet l'erreur incorpore une information sur la variable latente ce qui explique aisément qu'une erreur (à un certain seuil pour ne pas tout bruite) puisse

« améliorer » la mesure. Sijtsma (2012b) rappelle que pour le modèle de Rasch « *only variation in the latent variable determines the probability of a particular score on item j; no other influences are active* » (p. 794). D'ailleurs Scholten (2011a) écrit « *Although the effect of error on precision is probably only favorable in very few cases* » (p. 64).

Sijtsma (2012b) apporte une autre réponse qui me paraît bien plus pertinente. Certes, si le modèle de Rasch n'est que le modèle de Guttman additionné d'une erreur, cela ne permet pas de passer d'un cadre ordinal à un cadre quantitatif. Le modèle de Rasch considère que la différence entre l'habilité et la difficulté détermine une probabilité de réussite qui varie monotoniquement. Le modèle de Guttman ordonne les personnes alors que le modèle de Rasch donne une statistique suffisante pour estimer la variable latente sur une échelle métrique (sur la statistique suffisante, voir Leplège, 2003b). La différence entre l'habilité et la difficulté n'affecte pas le modèle de Guttman, c'est donc une information non intégrée, alors que le modèle de Rasch intègre cette information. La différence entre le modèle de Guttman et le modèle de Rasch n'est donc pas qu'une différence d'erreur de mesure, mais une différence de l'information prise en compte. Nous retrouvons en quelque sorte un axe argumentaire proche de celui de Scholten : il y a une information en plus dans le modèle de Rasch. Mais ici, l'information n'est pas liée à une erreur « informative » comme Scholten essaye de le montrer, l'information que le modèle de Rasch intègre est liée à la variable latente, donc sans contradiction avec le cadre formel de ce modèle (la différence des scores réels entraîne des probabilités différentes avec Rasch, alors qu'avec Guttman nous avons une variable dichotomique). Si nous reprenons l'argumentaire de Scholten, ce n'est pas l'erreur informative qui expliquerait l'efficacité du modèle de Rasch, mais le fait que la mesure du test intègre l'amplitude des différences de l'attribut latent.

Heene (2013) contredit Sijtsma « *I do not see why this line of argumentation refutes Michell's(2008) "Rasch paradox". Sijtsma's reasoning presupposes that the latent trait is continuous. Furthermore, we can only ignore information "...contained in the intervals" when there already is interval-level information, but thi is not at all self-evident but simply an assumption of IRT.* ». En retour, je crois que Sijtsma n'adhère pas à priori au modèle continu (cf. ses autres articles), mais il discute sur la base de la modélisation différenciée entre Guttman et Rasch. Je trouve que la réponse de Sijtsma est pertinente pour considérer que le modèle de Rasch ne souffre pas de ce paradoxe pointé par Michell.

2.1.7.3 Le modèle de Rasch est-il une mesure conjointe?

Remarquons que le paradoxe Guttman-Rasch pointé par Michell (2008a, 2009) rejoint dans ses fondements la critique de Kyngdon (2008b, 2013) : la concaténation de l'habilité et de la difficulté génère non pas un élément perceptible, mais une probabilité. C'est le caractère spécifique de la probabilité qui produit le paradoxe Guttman-Rasch discuté par Michell mais dont les fondements sont dans la discussion portée par Kyngdon. Pour ces auteurs, avoir une variable latente mesurable avec le modèle de Rasch nécessite que l'habilité et la difficulté soient mesurables. Comme nous l'avons vu, ces critiques peuvent elles-mêmes être critiquées. Pour autant, le modèle de Rasch pose certains problèmes qui paraissent incontournables.

1. *La précision de la mesure*

Si nous reprenons le point de vue « réaliste » (defendu par Michell, Sijtsma, Borsboom, Scholten mais non par Kyngdon), le modèle de Rasch pourrait répondre au cadre de l'ACM s'il n'y avait pas d'erreur de mesure et si le cadre théorique était clarifié. Sijtsma (2012b) rappelle qu'il faut distinguer le modèle théorique, les données collectées et le modèle statistique utilisé. Le modèle théorique n'a rien à voir avec le modèle statistique. La structure des données est déterminée par l'opérationnalisation du test. Le modèle statistique analyse les données. Le plus souvent, le modèle théorique est mal défini, donc les données peu précises et ne représentent pas l'attribut théorique. Le modèle statistique, comme celui de Rasch, ne peut détecter de telles incorrections. Les approches itératives (bootstrap par exemple) qui permettent d'estimer le meilleur modèle sont risquées car une procédure exploratoire ne permet pas de valider un modèle théorique. Sijtsma (2012b) donne une illustration en comparant différents résultats de la littérature académique sur un modèle de Rasch et pointe la difficulté de rapprocher un modèle empirique d'un modèle théorique.

2. *La fiabilité de la mesure*

Certains auteurs qui défendent le modèle de Rasch, Borsboom & Scholten (2008) par exemple, alertent sur la prudence avec ce modèle car il peut ne pas respecter les axiomes des mesures conjointes malgré des données apparemment fiables. Scholten (2011b) discute des problèmes avec le modèle de Rasch comme l'élimination d'items non conformes avec les axiomes des mesures conjointes permettent de construire un modèle apparemment fiable, mais sans comprendre les raisons de l'inadéquation de certains items et donc peut être une variable confondante liée au modèle.

Michell (2004) rappelle que des données simplement ordinales peuvent valider le modèle de Rasch, c'est-à-dire que des modèles non-quantitatifs valident l'IRT, ce qui permet de confirmer que la validation du caractère quantitatif n'est pas testable de cette sorte.

3. *La difficulté à définir la difficulté*

Trendler (2013) cite Hambleton, Swaminathan, and Rogers (1991) qui discutent de l'unidimensionnalité de l'habileté de façon critique. : « *These factors might include level of motivation, test anxiety, ability to work quickly, tendency to guess when in doubt about answers, and cognitive skills in addition to the dominant one measured by the set of test items* ». De plus, comme le discute Scholten (2011a), il est difficile d'obtenir une infinité de « difficultés » à la différence d'objets physiques car la différence de difficultés est difficile à établir, or les axiomes des mesures conjointes additives imposent cette contrainte.

2.1.7.4 Le modèle de RASCH est-il efficient ?

1. *Le modèle*

Van der Maas et al. (2011) rappellent que le modèle de Rasch est basé sur un choix mathématique non argumenté dans le cadre des théories psychologiques : « *In introductions to IRT, the preference for the logistic equation is typically explained in terms of statistical or measurement theoretical convenience. However, from a substantive point of view, the lack of a psychological justification for this key property of the measurement model compromises test validity.* » (p. 353). Heene (2013) souligne la même chose : « *It is mostly forgotten that Rasch himself did not derive his model from empirical observations but “. . .within [Rasch's] own mathematical play ground—with no relation to any actual item analysis problem!” (Rasch,1979). It is not necessarily wrong to develop mathematical models independently from empirical observations. But, it is also not at all *elf-evident* that empirical insights will result from such models* » (p. 247). Leplège (2003b) apporte des éléments historiques pour montrer que le modèle de Rasch a été construit sur la base initialement de la loi de Poisson qui semblait correspondre au cadre probabiliste des erreurs, donc avec une justification théorique du modèle. Cependant, le texte de Leplège laisse considérer que ce sont in fine les qualités mathématiques du modèle probabiliste retenu (le modèle de Poisson multiplicatif) qui semblent avoir privilégié son choix, et non un cadre théorique justifiant ce modèle probabiliste.

2. L'efficience du modèle

Van der Maas et al. (2011) développent un modèle (« Q-diffusion ») qui intègre le cadre axiomatique du test et de sa mesure en cadrant ce qu'est l'habilité. Leur résultat montre qu'un test semble non-fonctionnel avec le modèle IRT, mais fonctionnel pour des mesures d'attitude : « *This is surprising because IRT models have been traditionally proposed for, and applied in the context of, ability testing* » (p. 353). Ce résultat obtenu par des défenseurs du modèles de Rasch (Borsboom entre autres) questionne donc l'efficience de ce modèle dans un cadre théorique clairement identifié de ce qu'un test est sensé mesuré (l'habilité). P. Barrett (2005) revient lui aussi sur la fragilité du modèle de Rasch à partir d'illustrations empiriques qui marquent sa faiblesse relativement à ses hypothèses.

2.1.7.5 Conclusion sur le modèle de RASCH

En conclusion, sans pouvoir affirmer à l'encontre de tout un courant voyant dans le modèle de Rasch le seul modèle avec variables latentes vérifiant le cadre formel de la mesure, un certain nombre de discussions convergent vers une remise en question des qualités de mesure du modèle de Rasch. La variable latente est la performance au test, et elle vérifie théoriquement les axiomes de l'ACM. Mais nous n'avons qu'un petit nombre d'items ce qui rend difficile l'assimilation à une fonction continue. De plus, l'erreur de mesure entraîne la violation des axiomes de l'ACM. Enfin, les questionnements récents du modèle de Rasch dans le cadre d'une théorie substantive d'un test (Van der Maas et al., 2011) renvoie à son efficience réelle. In fine, la littérature académique récente semble converger sur la remise en cause du modèle de Rasch comme modèle de mesure répondant au cadre formel de la mesure en physique, d'où le développement de la *latent theory* par Markus et Borsboom (2013) pour essayer de donner un autre cadre formel de la mesure à la psychologie.

2.1.8 Conclusion

La première approche pour valider la mesurabilité d'un attribut est de valider expérimentalement la nature additive de l'attribut lorsque la mesure est extensive. Pour Michell (1999), cette vérification n'est jamais faite. Pour Trendler (2009), cette vérification est impossible en psychologie à l'image de ce que fait la physique pour valider les mesures d'objets dont la mesure serait intensive. En effet, pour vérifier empiriquement l'hypothèse quantitative d'un objet dont la mesure serait intensive (pas directement observable et non-

directement testable), il faut le tester indirectement. Il faut donc faire varier en intensité les paramètres et observer les résultats empiriquement. Une propriété mentale n'est donc mesurable que si elle est dépendante de conditions manipulables. Manipuler des propriétés mentales est théoriquement faisable, mais il paraît impossible que l'on puisse manipuler le degré requis pour valider l'hypothèse quantitative. Il faudrait arriver à utiliser des variables concomitantes pour valider cela et pouvoir contrôler les erreurs systématiques (l'attention des sujets, ...). Pour Trendler (2009), les propriétés mentales ne sont pas assez manipulables et contrôlables (au sens expérimental) pour valider empiriquement leur caractère quantitatif.

La seconde méthode en dehors de l'approche formelle est l'ACM pour vérifier la mesurabilité d'un construit. Il faut avoir un bi-ordre (habileté et motivation par exemple) et son association donne le test. Or, comme nous l'avons vu, deux problèmes existent en psychologie avec l'ACM : 1/ les données en psychologie intègrent une erreur aléatoire et une erreur systématique non-estimables, qui amènent à violer les axiomes de l'ACM, l'ACM demande ainsi trop de précision pour être utilisable en psychologie (Sijtsma, 2012b) ; 2/ il y a une impossibilité pratique à vérifier tous les axiomes de l'ACM.

L'approche « statistique » (dite *Latent variable theory*, par Markus et Borsboom) propose un cadre formel entre le cadre classique et la RTM. Cette approche se veut inscrite dans l'empirisme-réaliste, mais ne discute pas d'un cadre formel de validation d'une mesure. Or il faut discuter de ce que l'on veut mesurer avant de le faire. Il faut nécessairement partir de la propriété mentale, et valider sa mesurabilité pour poser ensuite un modèle statistique, et non se contenter comme Markus et Borsboom de valider la mesure par le modèle statistique.

Le modèle de Rasch souffre des limites du modèle d'analyse conjointe, avec, de plus, des problèmes spécifiques que j'ai développés précédemment (la difficulté de poser la difficulté entre autres), qui fragilisent sa prétention à répondre au cadre formel de la mesure.

Pour Sijtsma (2012b), demander à la psychologie de se référer à la physique comme le demande entre autres Kyngdon (2008b) et Michell (2008a), et obtenir une précision identique, est un but inatteignable. Et de fait, il n'y a pas de protocole formel possible pour valider le caractère mesurable d'une propriété mentale. Même si certains auteurs, comme Heene (2013), semblent considérer comme aisée une vérification de la nature mesurable d'un construit psychologique, ceci entre en contradiction avec toute la littérature : « *An attribute possessing a quantitative structure is required to satisfy the three conditions of ordinality (transitivity, anti-symmetry, and strong connexity) and the six conditions of additivity (associativity, commutativity, monotonicity, solvability, positivity, and the Archimedean condition; cf. Michell, 1990, p.52f.). Most of these conditions are testable hypotheses but I have never seen*

any empirical test in psychological articles before data were analyzed with SEM or IRT models, which already assume the quantitative structure of the attributes under consideration as argued below. ». Certes, personne ne semble vérifier la mesurabilité d'une propriété mentale, mais cela n'est pas possible selon le cadre formel de la mesure édictée par elle.

En conséquence, certains préconisent le recours à des mesures simplement ordinales. Pour Sijtsma (2012b), le manque de précision, les erreurs aléatoires et systématiques, font que la mesure en psychologie ne peut être précise et doit accepter des formes faibles de mesure : ordinales ou nominales. De même, Michell (2004) conseille les modèles ordinaux plutôt que quantitatifs.

A l'inverse, d'autres auteurs sont pour conserver des approches quantitatives quand cela paraît cohérent. Pour Markus & Borsboom (2012), le problème est de savoir si nous avons une mesure d'un attribut quantitatif, même si nous ne pouvons pas le prouver (illustrant cela par le fait que les égyptiens ont effectué des mesures pour construire les pyramides sans utiliser la théorie de la mesure arrivée par la suite). Pour eux, un siècle de tests montrent que nous avons des résultats stables, prouvant qu'il y a quelque chose au-delà des données, et que les données ne sont pas aléatoires. Borsboom & Mellenbergh (2004) considèrent de même que la démarche avec les variables latentes est scientifique, en hypothésant que les attributs sont quantitatifs, et en vérifiant sur la base de données cette hypothèse. Cependant, la validation scientifique est compliquée. Borsboom & Mellenbergh (2004) reconnaissent : « *it is certainly true that IRT models do not permit a direct test of the hypothesis that an attribute is quantitative* » (p. 113). Mais le point de vue « statistique » de Borsboom & Mellenbergh (2004); Borsboom & Scholten (2008); Markus & Borsboom (2013a) renvoie pour la littérature académique au fait que les résultats statistiques ne permettent pas, le plus souvent, de trancher entre une propriété mentale purement ordinale ou quantitative (Sijtsma, 2012b). Trendler (2013) rappelle que ce n'est pas parce que les données valident empiriquement un modèle que ce dernier est correct pour représenter une propriété, rappelant les différentes données estimables par un modèle de Rasch sans correspondre à ce modèle : « *in reality psychological attributes might be quantitative, but if they really are we do not know, and we never will know.* » (p. 610).

Si nous partons d'un point de vue réaliste de la mesure, comme le pose Michell (la valeur d'un rapport entre deux attributs, qui caractérise le rapport d'une qualité entre ces attributs, est un fait objectif synthétisé par une valeur, comme par exemple le rapport de 1 à 2 entre deux bouts de bois dont le second est d'une taille double du premier), la mesure devient inutilisable avec des propriétés mentales, car une telle approche réclame le recours au cadre formel de la

mesure. Markus & Borsboom (2013a) rappellent que « *No psychological concatenation operations existed, and hence nos psychological measurement was fundamental in Campbell's sense* » (p. 27), et dans ce sens, il ne peut y avoir d'équivalence entre la "propriété" mesurée et la "mesure" (p. 28-29-30). Pour McGrath (2005a, p. 116), « *Even if it is possible to provide an objectively defensible model of a complex construct, precise measurement of such constructs may not be possible.* ». McGrane (2015) pose les mêmes limites de la psychologie, et appelle la psychologie à se rapprocher de la démarche scientifique de la physique : « *Quantitative theory and laws, as well as the system of physical equations and units that they determine, are the actual foundations of scientific measurement* » (p. 432). La psychologie semble donc dans une impasse concernant l'usage de la mesure : vouloir utiliser les mesures et se confronter à l'impossibilité de valider cette démarche.

Je considère que cette discussion, aussi pertinente qu'elle soit, mérite un regard critique. Toute la discussion se base, plus ou moins implicitement, dans le cadre de l'empirisme-réaliste, c'est-à-dire dans un cadre où l'enjeu est de mesurer un objet fixe et extérieur au chercheur. En repositionnant l'objet de la mesure, à savoir les propriétés mentales, je considère qu'une démarche empirique de validation d'une mesure est possible. C'est l'objet de la suite de mon texte.

2.2 Mesurer quoi ?

Si nous partons d'un point de vue réaliste de la mesure, comme le pose Michell, la mesure devient inutilisable avec des propriétés mentales car une telle approche réclame le recours au cadre formel de la mesure en physique, ce qui est impossible en psychologie (Trendler, 2013). Le cœur de ma critique sur la littérature académique est sur l'objet questionné par une mesure. Nous discutons de la mesure de propriétés mentales. Les propriétés mentales sont des objets réels mais à appréhender par le prisme de la pratique, leur nature ontologique est différente des objets que la physique mesure. Il faut donc un protocole de validation de la mesure spécifique au champ de la psychologie, en rupture avec le cadre formel de la mesure en physique. Je m'inscris donc en opposition aux auteurs voulant faire entrer la validation formelle d'une qualité mesurable en psychologie dans le cadre formel de la mesure en physique (McGrane, 2015; Michell, 2008b; Salzberger, 2013). Pour mener cette discussion je vais repartir de la nature ontologique des propriétés mentales déjà discutée, pour glisser vers un cadre de validation de la mesure.

2.2.1 L'objectivité de la propriété mentale ?

Dans cette section je reviens sur les propriétés mentales et leur nature ontologique pour préciser ensuite plus clairement comment leur mesure peut s'envisager. Cette partie est donc le prolongement de la discussion épistémologique abordée précédemment. Si je me permets ce développement maintenant c'est que cela me paraît impératif de repositionner ce que nous cherchons à mesurer avec des propriétés mentales. Par ailleurs, mon objectif n'est pas de prendre parti dans des discussions contradictoires en philosophie, cependant je pointerai une divergence entre Putnam et Searle qui me paraît nécessaire pour la suite.

2.2.1.1 La connaissance des propriétés mentales

Putnam (1984) pose la question : « *A l'extérieur il y a des objets. De l'autre côté il y a l'esprit-cerveau qui pense ou qui traite de l'information. Comment les signes du penseur établissent-ils une relation de correspondance unique avec des objets ou des ensembles d'objets extérieurs ?* » (p. 63). Putnam pose un cadre en rupture avec le relativisme et l'empirisme-réaliste car les catégories pensées par l'homme sont en tension entre réalisme et subjectivité. Putnam affirme vouloir « *briser l'emprise étouffante qu'un certain nombre de*

dichotomies semblent exercer sur la pensée [...] La plus importante d'entre elles est la dichotomie entre les conceptions objectives et les conceptions subjectives de la vérité et de la raison » (p. 7). Putnam s'efforce « *de développer une conception de la vérité qui réunit les composantes subjectives et objectives* » (p.8). Pour Putnam, « *L'esprit peut saisir des choses qui lui sont extérieures* » (p. 39), donc en opposition avec tout un courant relativiste ; et pour lui c'est l'efficacité de nos intentions dans nos actions qui valide nos théories « *il est justifié de penser que notre théorie du monde ordinaire est au moins à peu près vraie et que nous ne pourrions pas survivre si ce n'était le cas* » (p.52). Nous retrouvons en quelque sorte le célèbre proverbe anglais « *the proof of pudding is in the eating* »²⁴. Pour Putnam (1984) « *Le seul critère pour être un fait, c'est d'être rationnellement acceptable [...] l'acceptabilité rationnelle et la vérité sont deux notions reliées mais distinctes* ». Putnam s'accorde avec le fait « *qu'il n'existe pas de canon fixe et a-historique qui définisse la rationalité* ». Putnam développe ainsi une vision de la pensée humaine de façon interactive : il faut des « *conditions préalables de la représentation, de la pensée, de la référence [...] on ne peut pas faire référence à certaines choses si on n'a aucune interaction causale avec elles* » (p. 27). Putnam pose un point de vue « *internaliste* ». Il considère être en rupture avec la conception d'un monde « *constitué d'un ensemble fixe d'objets indépendants de l'esprit* » pour lequel « *Il n'existe qu'une seule description vraie de "comment est fait le monde"*. Putnam pose : « *La "vérité" est une sorte d'acceptabilité rationnelle (idéalisée) – une sorte de cohérence idéale de nos croyances entre elles et avec nos expériences telles qu'elles sont représentées dans notre système de croyance – et non une correspondance avec des "états de choses" indépendants de l'esprit ou du discours. Il n'y a pas de point de vue de Dieu qui soit connaissable ou utilement imaginable ; il y n'a que des différents points de vue de différentes personnes qui reflètent les intérêts et objectifs de leurs descriptions et théories* » (p. 61-62). Putnam (2011) parle d'un réalisme avec un petit « *r* » et s'oppose à considérer que notre conception commune du monde comporte une image fautive de la réalité « *notre image du monde ne peut être "justifiée" par rien d'autre que par son succès* » (p. 145), mais sans voir une démarche instrumentaliste. « *La vérité n'est rien de plus que l'acceptabilité rationnelle idéalisée [...] Tout ce que je demande, c'est que ce qui est censé être "vrai" puisse être garanti sur la base de l'expérience et de l'intelligence à des êtres dotés d'une "nature rationnelle et sensible"* » (p. 162).

²⁴ il semble difficile de donner une réelle paternité à cette phrase : <http://www.phrases.org.uk/meanings/proof-of-the-pudding.html>

Nous retrouvons dans cette position ce que j'ai appelé précédemment le pragmatisme-réaliste (que Tiercelin, 2002, discutant de Putnam, appelle *réalisme pragmatique*), considérant que l'esprit peut objectiver par le langage la réalité, mais en considérant cette objectivité comme une interaction, « *l'objectivité-pour-nous* » (Putnam, 1984, p.64) par la praxis concrète : « *reconnaitre que mes jugements prétendent à la validité objective et reconnaître que je suis façonné par une culture particulière et que je parle dans un contexte particulier ne sont pas choses incompatibles [...] Comme Dewey et Peirce nous l'ont appris, les vraies questions doivent avoir un contexte et un enjeu précis [...] cela veut dire que l'on renonce à une certaine image métaphysique de l'objectivité, mais cela ne veut pas dire que l'on renonce à penser qu'existent ce que Dewey appelait des "résolutions objectives de situations problématiques" - des résolutions objectives à des problèmes qui sont situés en temps et lieu, par opposition à une réponse "absolue" à des questions "indépendantes de toute perspective". Et cela suffit pour qu'il y ait de l'objectivité* » (Putnam, 2011, p. 351). Pour Putnam « *Le fait de rejeter qu'il existe [...] une théorie qui serait simplement "vraie en soi" indépendamment des observateurs possibles ne revient pas à identifier la vérité avec l'acceptabilité rationnelle [...] la vérité est censée être une propriété inaliénable des propositions [...] la vérité est une idéalisation de l'acceptabilité rationnelle* » (Putnam, 1984).

Je me positionne relativement à la psychologie dans cette perspective de la connaissance ainsi posée par Putnam. Les propriétés mentales existent sous la forme de leur manifestation, donc sous la forme de l'interaction sociale. Ce sont des réalités objectives, en considérant cette objectivité comme construite par la fonction de ces propriétés mentales dans l'interaction sociale, dans « *un contexte et un enjeu précis* ». De plus les propriétés mentales sont des « *objectivités mouvantes* » comme je l'ai déjà écrit, différentes d'autres objets « *naturels* » et « *fixes* » extérieurs à l'homme. C'est bien par une praxis que nous avons une objectivation des propriétés mentales, une « *réalité pour nous* » où « *reconnaitre que mes jugements prétendent à la validité objective et reconnaître que je suis façonné par une culture particulière et que je parle dans un contexte particulier ne sont pas choses incompatibles* ». Je le redis, tout ce développement de Putnam est un développement généraliste, mais je ne me l'approprie que pour la psychologie sans me positionner sur d'autres champs scientifiques. Je considère que le cadre donné par Putnam est le cadre opérant pour appréhender notre connaissance des propriétés mentales, c'est-à-dire considérer l'objectivité des propriétés mentales en tant que pratique intersubjective.

2.2.1.2 Convergence avec Searle

Je m'appuie sur Putnam pour mon développement. Cependant, ce qu'écrit Searle (1996, 2000) me paraît en phase avec ce que je défends comme position. Searle (1996) pose « *Toute représentation a une forme aspectuelle. Elle représente sa cible sous certains aspects et pas sous d'autres. Bref, c'est seulement à partir d'un certain point de vue que nous représentons la réalité* ». Ceci me semble proche de ce que Putnam écrit car nous retrouvons en quelque sorte la « *réalité pour nous* » de Putnam. Searle développe par ailleurs toute une discussion sur l'intentionnalité, et le fait de devoir comprendre les états mentaux au-delà du cerveau individuel, qui me semble parfaitement en phase avec mon point de vue : « *Le mental est simplement un ensemble de caractéristiques physiques du cerveau à un niveau de description supérieur à celui des neurones [...] les états mentaux se distinguent des autres phénomènes physiques en ce qu'ils sont ou bien conscients ou potentiellement tels. Là où il n'y a pas d'accessibilité à la conscience, du moins en principe, il n'y a pas d'états mentaux* ». Je crois que ce que j'ai développé renvoie à ce point de vue de Searle en considérant les propriétés mentales comme une réalité intersubjective à partir de manifestations perceptibles révélant une propriété émergente (donc conscientisée ou potentiellement pouvant l'être).

Searle définit par ailleurs une entité *ontologiquement subjective* comme celle dont le mode d'existence dépend de ce qu'elle soit éprouvée par des sujets ; de nouveau je retrouve ce que j'ai développé comme objectivité subjectivée. Pour ce faire, Searle a recours au concept « d'arrière-plan » qui détermine un « *ensemble d'aptitudes, de tendances, de dispositions, et de capacités [...] sans lesquelles il ne lui serait pas possible de fonctionner, d'avoir des conditions de satisfactions. L'arrière-plan ne fait pas partie du sens et de l'intentionnalité, pourtant, sens et intentionnalité n'existent en tant que tels que par rapport à lui* » (Searle, 2000). Cet « arrière-plan » se rapproche à mon sens de la « matrice » de Hacking (2008, p. 26) : « *Les idées n'existent pas dans le vide. Elles existent à l'intérieur d'un contexte social. Appelons matrice ce à l'intérieur de quoi l'idée, un concept ou une espèce se forme* ».

2.2.1.3 Divergence avec Searle

Au final ce que Searle développe ne me paraît pas contradictoire avec le pragmatisme-réaliste que j'ai développé en m'appuyant sur Putnam. Mais il y a cependant une contradiction entre Searle et Putnam sur l'ontologie du réel. Même si je ne crois pas avoir nécessairement à prendre parti entre eux deux d'une manière globale sur le rapport « réalité »/« sens de la réalité », la manière dont je comprends cette discussion entre Searle et Putnam n'est peut-être

pas sans importance pour la suite de mon travail centré sur les propriétés mentales. Je ne me positionne que relativement aux propriétés mentales et mon objectif n'est pas d'épouser sans critique l'approche de Putnam, mais ce que propose Putnam me paraît devoir être une manière de comprendre notre objectivation des propriétés mentales plus efficiente que Searle.

Searle considère que « *la réalité ontologiquement objective n'a pas, quant à elle, de point de vue* ». Putnam réfute au contraire de pouvoir poser le réel indépendamment d'une représentation du réel. En effet, dire que « la réalité » n'a pas de point de vue, c'est considérer la réalité en dehors d'une relation. En tant que réaliste Putnam ne nie pas la « réalité » extérieure ni la capacité de l'objectiver (donc Putnam est en rupture avec le scepticisme relativiste), mais pour Putnam « *on ne peut pas faire référence à certaines choses si on n'a aucune interaction causale avec elles* ». Putnam considère que « *Ce que le réalisme externe [proposé par Searle] nous offre est un quelque chose qui est impensable, indescriptible, inaccessible, inconnaissable, inexprimable par les mots et, pour finir, dénué de sens. Le véritable problème d'un tel réalisme [externaliste] n'est pas d'être faux ; c'est, en dernier ressort, d'être inintelligible* ». Searle (1996) critique ce passage et affirme lui défendre un réalisme externaliste. Searle (2000) s'affirme par ailleurs « *internaliste* » ; donc au-delà des catégories en « isme » ce qui m'intéresse c'est le nœud entre Putnam et Searle sur le « réel » pour clarifier mon point de vue sur les propriétés mentales.

Je centre en effet ma discussion uniquement sur les propriétés mentales, et je devine donc entre Putnam et Searle une ligne de fracture sur ce point. Searle considère que « *La subjectivité ontologique de la réalité socialement construite requiert une réalité ontologiquement objective à partir de laquelle elle est construite* » et donc les « *états mentaux* » sont « *indépendants des représentations, mais pas de l'esprit* ». Je suis bien évidemment en accord avec le fait qu'il y ait une réalité des « propriétés mentales ». Mais pour Putnam, les états mentaux sont dépendants, dans leur subjectivisation, des représentations, c'est un point de vue réaliste-internaliste. Et sur ce point je m'accorde avec Putnam et Hacking (2008, p. 53) pour qui : « *Les classifications des sciences sociales sont interactives. Les classifications et les concepts des sciences naturelles ne le sont pas* ». Je ne veux pas entrer dans la discussion fondamentale de savoir si on ne peut considérer d'une façon générale la réalité qu'à partir du cadre conceptuel que nous avons pour les objectiver, c'est-à-dire de notre pratique avec le réel (ce que défend Putnam) ; ou au contraire considérer avec Searle que le réalisme est une théorie ontologique (l'ontologie objective), ce que semble aussi défendre Hacking pour les objets non-sociaux. Mais l'ontologie des propriétés mentales ne peut être pour moi qu'une ontologie subjective ; c'est-à-dire que notre manière de donner

du sens à ces propriétés mentales est autant réelle que construite. On ne peut séparer les deux. Il n'y a donc pas de « propriétés mentales » « *indépendantes des représentations* », pas de « *ontologie objective* » au sens de Searle pour les propriétés mentales, mais sans pour autant nier leur réalité.

Le point de divergence entre Putnam, Searle et Hacking qui dépasse le cadre de cette thèse n'entraîne pas de divergences entre eux sur le fait que les propriétés mentales ne puissent se résumer aux neurones, que ce sont des réalités objectivées intersubjectivement et donc devant être appréhendées comme telles (« l'ontologie subjective » avec « l'arrière-plan » de Searle, ou la « réalité pour nous » de Putnam, ou « la matrice » et « l'interaction » de Hacking). Mais où la divergence entre Putnam (et Hacking) et Searle me paraît importante c'est uniquement sur le problème de la catégorisation des propriétés mentales, et donc la nature ontologique de ces propriétés mentales. Pour Searle les propriétés mentales semblent pouvoir exister en tant que catégories indépendamment des représentations, c'est-à-dire que notre catégorisation de ces propriétés mentales pourrait avoir une réalité ontologique indépendante de nos pratiques. Or je considère, en accord avec Putnam et Hacking, que cette catégorisation des propriétés mentales est un processus objectivation intersubjectivé, « *autant construit que découvert* », et non une réalité ontologique indépendante du sens que nous lui donnons.

2.2.2 Catégorisation de la propriété mentale ?

Considérer la réalité comme une pratique oblige à discuter de la façon dont le cerveau objective/catégorise la réalité, c'est-à-dire comment nous donnons du sens au réel. Et c'est bien d'un point de vue réaliste-internaliste comme Putnam le pose que je me positionne et non sur une *ontologie objective*, comme je le comprends chez Searle, laissant considérer nos catégories comme potentiellement existantes indépendamment d'une pratique.

2.2.2.1 La catégorisation objective et subjective

En effet, la réalité apparaît de prime abord comme une unité et le cerveau semble devoir découper cette unité en catégories pour pouvoir objectiver le monde, lui donner du sens. La catégorisation est un processus d'objectivation de la réalité qui semble ainsi incontournable pour l'homme, Falissard (2012) « *Le processus de catégorisation semble inhérent au fonctionnement du cerveau* ». Comme le pose Pereira (2009) « *Le découpage de la réalité est ambigu. Il semble à la fois relever du fait, donc de la réalité, et du langage, donc des valeurs culturelles* » (p. 42). Ce processus de catégorisation, nécessaire pour penser le monde, est

donc un processus issu de la praxis de la personne, donc ni « vrai » ni « faux » mais relativement au processus d'apprentissage, au contexte/enjeux/matrice/arrière-plan (au sens de Hacking ou Searle donnés précédemment) que je résume par « pratique » ou « praxis » car « *autant construits que découverts* ». Putnam considère en effet « *Les «objets» n'existent pas indépendamment des cadres conceptuels. C'est nous qui découpons le monde en objets lorsque nous introduisons tel ou tel cadre descriptif [...] les objets sont autant construits que découverts, s'ils ne sont autant le fruit de notre intervention conceptuelle que le produit de la composante objective de l'expérience qui est indépendante de notre volonté, alors il est certain que les objets doivent se retrouver automatiquement sous certaines étiquettes, parce que ces étiquettes sont les outils que nous avons utilisés au départ pour construire une version du monde contenant ces objets* ». (p. 64) et d'ajouter « *C'est la pratique qui détermine l'interprétation [...] L'interprétation de la théorie psychologique et de sa terminologie est déterminée par notre pratique réelle [...] il est vain d'essayer de se représenter ce que sont réellement les faits perceptuels, indépendamment de la manière dont nous les conceptualisons [...] Les propriétés des choses interviennent effectivement dans des explications causales* » (p. 79-80-81).

Discutant de catégories de propriétés mentales, je ne généralise pas la discussion et me centre sur ces propriétés mentales car comme Hacking (2008) le pose, pour les objets « naturels » (non inscrits dans des interactions sociales), la tension entre « ontologie objective » vs « ontologie subjective » est questionnable. Pour Quéré (2001) les pragmatistes « *n'envisageaient donc pas de solution de continuité entre les sciences naturelles et les sciences humaines* » (p. 285). Les pragmatistes ne suppriment pas les différences catégorielles, mais dans une démarche « *empirique et "expérientielle"* » pour les sciences sociales, donc en ruptures avec des « *catégories naturelles* ». Des catégories à « *réadapter [...] en réajustant le cadre catégoriel* » (Quéré, 2001). Ainsi les catégories pensées par l'homme pour les propriétés mentales sont des catégories nécessaires, nécessaires pour conceptualiser le monde, mais aussi nécessaires relativement à notre praxis. Le langage est « *un instrument, étroitement associé à des actes, dans des circonstances qui appartiennent à notre vie* » (Cometti, 2010, p. 211). La catégorisation des propriétés mentales n'est pas une catégorisation d'objets externes à l'homme, mais un processus de catégorisation objectivé par notre pratique ; et c'est là où je me situe en contradiction avec Searle et en accord avec Putnam et Hacking.

Nous cherchons à regrouper dans une même catégorie des manifestations pratiques similaires. « *Nous sommes prêts à considérer que quelque chose appartient à une espèce [...] si cette*

chose possède la même nature essentielle, ou plus vaguement si elle est suffisamment "semblable" aux cas paradigmatiques [...] ce qui constitue une similitude suffisante est déterminée à la fois par l'espèce naturelle et le contexte » (Putnam, 1984, p. 119).

2.2.2.2 Une démarche pragmatique

Une telle position pragmatique, liant la théorie et la pratique comme un seul et même processus, est orthogonale avec l'approche empirique-réaliste sur les catégories utilisées pour définir les propriétés mentales ; et c'est la raison de ma discussion contradictoire entre Searle et Putnam. L'objectivation d'une propriété mentale est une pratique sociale, liées à des manifestations considérées comme relevant d'une seule et même catégorie. C'est la pratique sociale qui nous pousse à modifier nos catégories. Il n'y a pas en miroir de nos représentations une réalité fixe que nous chercherions à caractériser, mais c'est la pratique des propriétés mentales qui donnent leur réalité « pour nous ».

Concrètement, comme je l'ai déjà dit, il n'y a pas en amont une réalité biologique fixe d'une propriété mentale qui permettrait de caractériser une propriété mentale comme une structure neurologique particulière, identique pour toutes les personnes ayant cette même propriété mentale. Cela ne veut pas dire que certaines structurations neurologiques ne puissent pas déterminer certaines propriétés mentales, mais d'une façon globale on ne peut caractériser une propriété mentale par un fait biologique précis. Je le répète, cela ne renvoie pas à une position relativiste ou constructiviste radical car il y a une objectivité matérielle de la propriété mentale. Il y a bien une structuration neuronale, mais en mouvement car à comprendre dans l'interaction du sujet avec son environnement. Ce mouvement n'interdit pas une forme de stabilité, mais cette stabilité de manifestations n'équivaut pas forcément à une structuration statique de l'être biologique (une même propriété mentale est multiréalisable). La catégorisation n'est qu'un processus d'objectivation de cette réalité... mais catégorisation d'une « réalité pour nous » (au sens social du terme), d'une réalité qui ne peut être appréhendée que par ses manifestations. On ne peut pas considérer une catégorie de propriété mentale comme une réalité figée et externe à notre vie sociale, ni comme une simple construction indépendante de la réalité. Il faut renégocier en permanence les catégories posées des propriétés mentales, non que les anciennes catégories soient « vraies » ou « fausses » de façon a-historique, mais parce que ce sont toujours des catégories liées à une pratique et susceptibles d'évoluer, donc objectivées au moment de la catégorisation. Mais ce processus de renégociation de l'objectivation de la réalité n'est pas à considérer comme une

« construction » de la réalité sans soubassements objectifs. Je renvoie aux exemples développés par (Hacking, 2008, Chapter 4) sur « le faible d'esprit », « l'autisme infantile », « la schizophrénie », « l'anorexie », et les renégociations de ces pathologies dans l'histoire, la tension entre « biologie » et « construction sociale », pour illustrer la « *tension entre le réel et le construit* » en psychologie.

Il n'y a donc jamais de catégories figées avec les propriétés mentales, donc pas de propriétés mentales qui soient *ad vitam aeternam* « vraies ». C'est une des sources d'erreur primordiale à mes yeux, issue d'une épistémologie empirique-réaliste, que de considérer avoir construit une formalisation « fiable » ou « pérenne » d'une propriété mentale par variables latentes (un test qui serait valide une fois pour toute pour mesurer une propriété mentale). Cette formalisation ne peut se faire que construite dans l'interaction pratique de son usage. Cela ne veut pas dire repartir de zéro à chaque fois, mais cela veut dire ne pas considérer que la catégorisation et formalisation anciennes puissent être utilisées sans regard critique. « *Le monde empirique [...] dépend de nos critères d'acceptabilité rationnelle. Nos critères d'acceptabilité rationnelle nous aident à construire un tableau théorique du "monde empirique" et, à mesure que se développe ce tableau, il nous conduit à réviser nos critères d'acceptabilité rationnelle, et ainsi de suite à l'infini* » (Putnam, 1984, p. 152).

2.2.3 Unidimensionalité de la propriété mentale ?

Un point central de la discussion sur les variables latentes est « l'unidimensionnalité » de la propriété mentale. Cela va me servir comme levier critique sur la manière dont la psychologie aborde les propriétés mentales d'une façon commune.

2.2.3.1 Unidimensionalité de la propriété

Il est nécessaire que la propriété mesurée soit sur un domaine unique et non une propriété renvoyant à différentes catégories. Le problème central posé par des auteurs est la granularité du phénomène étudié (Mislevy, 2009), lié au problème du domaine unique des tâches demandées lors d'un test. Pour Michell (2012) les tests utilisés en psychologie sur des différences de difficultés, et les ressources cognitives utilisées pour répondre aux questions, sont certainement de natures différentes, et donc révèlent des domaines différents. L'intelligence requiert par exemple au moins habilité spatiale, habilité logique et habilité verbale, donc il n'y a pas une propriété unique (Markus & Borsboom, 2013a). L'unidimensionnalité questionnée peut s'étendre aux mesures d'attitude, pour exemple

l'anxiété est questionnée comme un domaine unique (P. Barrett, 2005) ou la dépression (Markus & Borsboom, 2013a).

Cependant, j'ai un regard critique sur cette manière de poser la discussion. La question du « domaine unique » ne doit pas se baser sur les processus neuronaux en amont qui font émerger la propriété mentale. Comme déjà discuté précédemment, il y a à mes yeux une erreur fondamentale de vouloir relier une propriété mentale émergente et les processus mentaux la générant, car nous ne discutons pas au même niveau de la réalité. Le « domaine unique » des manifestations d'une propriété mentale oblige à raisonner au niveau de la propriété mentale. La question est donc non pas de savoir si les processus en amont sont d'un domaine unique, ce qui n'est jamais le cas, mais de savoir si la propriété émergente à une réalité pratique unidimensionnelle. Il faut donc partir de la propriété se manifestant et non de ce qui la provoque pour questionner le domaine unique.

2.2.3.2 Une démarche pratique

Pour questionner ce domaine unique, comme dit précédemment, c'est une approche erronée que de considérer pouvoir conceptualiser une propriété comme unidimensionnelle sans poser la pratique liée (contexte et enjeu). En effet, la littérature abonde de propriétés considérées auparavant comme unidimensionnelles et qui par la suite sont appréhendées comme relevant de différents domaines. L'unidimensionnalité et le domaine unique n'existent pas en soi, c'est une catégorisation pratique, donc réelle et relative à la fois, et en conséquence cela explique les renégociations permanentes de ces catégories dans la littérature. Toute discussion sur le domaine unique de la propriété mentale (et non des processus en amont) doit donc questionner la pratique de cette propriété mentale.

Ma position est donc critique des travaux essayant de conceptualiser de façon statique une propriété mentale comme un objet réel extérieur à nos pratiques sociales. Nous pouvons accepter une conceptualisation d'une propriété mentale comme unidimensionnelle si on considère cette catégorisation comme renégociable, car inscrite dans un contexte et des enjeux précis. Le problème fondamental est de savoir s'il y a une émergence unique, stable et suffisamment homogène d'une propriété mentale pour questionner le domaine unique relativement à notre praxis au moment de cette conceptualisation.

2.2.4 Indépendance locale ?

Markus & Borsboom (2013a) rappellent qu'il faut pour une variable latente que les indicateurs (les variables manifestes formalisant les manifestations) proviennent « *from homogenous domains items* ». Mais de plus, la nécessité d'avoir des indicateurs d'une propriété mentale répondant à l'indépendance locale est centrale pour modéliser une propriété mentale par une variable latente. Cette discussion est reliée à la question précédente (l'unidimensionnalité de la propriété mentale), mais je fais quand même un focus sur ce problème central de l'indépendance locale avec les modèles avec variables latentes car il permet de mieux formaliser ma position critique.

2.2.4.1 La granularité

Concrètement, pour répondre au cadre formel des variables latentes, les manifestations d'une propriété mentale ne doivent être liées entre elles que relativement à la propriété mentale étudiée, et indépendante entre elles autrement. Cela veut dire, pour simplifier, que les covariations entre deux manifestations ne doivent être causalement créées que par la propriété mentale questionnée. Cette nécessité renvoie à l'hypothèse de « clôture » du modèle : « *L'hypothèse de clôture présume que chaque variable non incluse dans le système affecte au plus une variable du système, ce qui implique des corrélations nulles entre les termes d'erreur associés à chaque variable du système. L'hypothèse de clôture ne présume, par contre, pas que toutes les variables susceptibles d'avoir un effet sur le phénomène que le système formel représente sont incluses dans le système* » (Juhel, 2015) . C'est à notre avis le point essentiel pour déterminer la granularité nécessaire pour valider la catégorisation d'une propriété mentale et donc la variable latente associée. Bien évidemment, si le phénomène étudié est issu d'un système dynamique complexe, il y a de fortes chances que les processus cognitifs utilisés en amont agissent sur d'autres phénomènes que celui étudié, et donc que cela contredise l'indépendance locale des processus cognitifs. Mais au niveau des propriétés mentales émergentes, cela doit être analysé indépendamment des processus en amonts.

La littérature considère qu'il n'y a pas de protocole de validation de l'indépendance locale (Markus & Borsboom, 2013a). Il faut donc questionner cette indépendance locale relativement à la théorie pour la plupart des auteurs, et questionner si les manifestations associées d'une propriété mentale pourraient être reliées à d'autres propriétés mentales. Discuter du rôle de la théorie, comme le propose la littérature, renvoie le plus souvent à une approche empiriste-réaliste, comme si la théorie pouvait nous permettre in fine d'être capable

de définir un objet réel fixe et indépendant de sa fonction interactive, et en conséquence comme si la littérature pouvait un jour poser des manifestations pérennes et définitives pour mesurer la propriété mentale. Comme si on pouvait un jour poser un cadre théorique définitif d'une propriété mentale et de ses manifestations.

En fait, questionner la théorie est ce processus concret de renégociation tel que posé par Putnam, non comme critique des conceptualisations passées, mais comme processus de conceptualisation et catégorisation *ici et maintenant*. Il faut partir de la catégorisation qui est dépendante de la pratique concrète (contexte et enjeu précis), poser le cadre théorique en relation avec notre pratique. La propriété mentale est un objet social réifié par une catégorie, elle est donc uniquement vraie relativement à notre praxis (au sens social). Les manifestations utilisées pour le modèle de mesure doivent être dans le même cadre.

2.2.4.2 La renégociation

Donc valider l'indépendance locale doit être en écho avec le processus de catégorisation. L'indépendance locale doit être appréhendée par la formalisation d'une variable latente circonscrite à une propriété mentale potentiellement comme cause unique conjointe de ses manifestations. Il n'y aura pas de « vérité », car, comme le dit Putnam, c'est une idéalisation ; mais il faudra objectiver cette indépendance locale par une rationalisation : est-ce que les manifestations de la propriété mentale peuvent être considérées comme localement indépendantes ? Cette indépendance locale n'est pas une « vérité » indépendante de notre pratique, c'est-à-dire de notre catégorisation des propriétés mentales et des enjeux associés. Il faut donc à mes yeux questionner les autres catégories posées par le cadre théorique (les autres propriétés mentales conceptualisées de façon robustes relativement à notre démarche de mesure), pour questionner si des manifestations de la propriété mentale questionnée peuvent être dépendantes d'autres propriétés mentales. Je le rappelle, cette indépendance locale ne doit pas questionner les processus en amont mais bien la propriété émergente. Dans le futur, une renégociation de cette propriété mentale et de ses manifestations, peut amener à réviser le cadre théorique posé (je renvoie de nouveau aux exemples développés par Hacking, 2008). Cela n'interdit pas de poser ce cadre dans la pratique concrète de la propriété mentale au moment de son analyse.

Concrètement, il faut questionner les manifestations utilisées. Peuvent-elles être reliées à d'autres propriétés mentales ? Si nous considérons que les manifestations utilisées sont causalement dépendantes d'une autre propriété mentale que celle questionnées,

l'indépendance locale sera violée relativement à notre cadre théorique. Si nous devons considérer que les manifestations violent l'indépendance locale, cela renvoie certainement à l'unicité de la propriété mentale questionnée car une zone de recouvrement avec une autre propriété mentale semble exister. Cette renégociation est donc nécessaire à partir du cadre théorique posé, c'est-à-dire à partir du cadre pratique au moment de la mesure.

C'est cette absence de renégociation qui me semble critiquable dans l'usage de mesure de propriété mentale car il n'y a pas de « théorie » en soi d'une propriété mentale. La propriété mentale est un objet social, donc sa théorisation est sociale. En cela je suis en parfait accord avec Sijtsma (2012b) qui considère que ce sont des effets d'inertie et de conformisme qui pilotent à tort la plupart des mesures en psychologie. Il faut donc affirmer rompre avec le conformisme et rompre avec l'affirmation qu'une mesure ait été « validée » pour une propriété mentale indépendamment de son contexte pratique. Toute procédure doit se renégocier, c'est-à-dire par une démarche théorique au sens de la renégociation de la théorie. Il en va de même pour l'indépendance locale.

2.2.4.3 L'exemple de la dépression

Par exemple poser une mesure de la dépression comme l'échelle de Hamilton doit questionner si d'autres propriétés mentales ne peuvent pas être considérées comme causalement liées aux manifestations intégrées dans l'échelle de Hamilton. Si tel était le cas, nous violons le cadre de l'indépendance locale des manifestations utilisées dans l'échelle de Hamilton, et donc en conséquence le cadre formel nécessaire à son usage. Je ne peux prétendre apporter de réponse car je ne suis pas expert de la dépression, mais je subodore que cette échelle de Hamilton, qui semble très utilisée, renvoie la question du « domaine unique » de la dépression telle que appréhendée aujourd'hui dans notre réalité sociale (comme le pose Markus & Borsboom, 2013a).

En effet, les items de cette échelle renvoient par exemple à « agitation » et « insomnie ». Ces manifestations me paraissent potentiellement pouvoir relever d'autres propriétés mentales que celle de la dépression, comme des symptômes de bipolarité telle que semble le définir la psychiatrie aujourd'hui. Donc l'échelle de Hamilton actuelle peut potentiellement souffrir de la non-indépendance locale des manifestations intégrées relativement au cadre théorique actuel. Elle doit être renégociée relativement au cadre pratique de son usage actuel, et relativement aux autres catégories psychologiques théorisées actuellement. On ne peut pas considérer que cette échelle ait posé de façon définitive la mesure de la dépression. Cette

discussion sur l'échelle de Hamilton sort du champ de mon travail et de mon domaine de compétence, mais illustre me semble-t-il l'inertie potentielle de l'usage d'échelle de mesure ou de tests alors que la renégociation doit être centrale.

Cette pratique nécessaire de renégociation renvoie à la difficulté de poser une mesure en psychologie car cela demande de questionner en permanence le domaine unique de la propriété mentale et l'indépendance locale des manifestations intégrées relativement au cadre actuel de nos catégorisations.

2.2.5 Conclusion

Dans cette partie je n'ai fait que repositionner la mesure dans son objectif en reprenant ce que j'ai déjà développé via le pragmatisme-réaliste. La propriété mentale mesurée ne doit pas être appréhendée comme une réalité indépendamment de sa pratique sociale. L'enfermement dans l'empirisme-réaliste et la quête d'une propriété mentale « vraie », conceptualisable par une « théorie » et en conséquence des « manifestations » répondant à l'indépendance locale de façon pérenne, est un leurre dans lequel sont enfermés bien des travaux de recherche. Les propriétés mentales sont des objets sociaux et la mesure doit donc être appréhendée de la sorte, ainsi la mesure doit partir d'une renégociation d'une façon pratique du domaine unique de la propriété mentale et de l'indépendance locale des manifestations. Cela ne veut pas dire glisser vers un constructivisme en considérant la propriété mentale comme une simple construction, comme je l'ai déjà dit. Il y a une objectivité, mais cette objectivité est inscrite dans une praxis. Il faut questionner l'objectivité de la propriété mentale, son unidimensionnalité, l'indépendance locale des manifestations, relativement à la praxis. C'est donc bien un processus d'objectivation d'une réalité, mais objectivité « pour nous ».

2.3 Comment mesurer pragmatiquement ?

Pour mener cette discussion je vais m'appuyer sur l'article de Sherry (2011), les travaux de Körner (1962, 1964). Je conserverai le terme de mesure au sens général, incluant les mesures ordinales, car, comme le disent Mari, Carbone, & Petri (2012) : « *Furthermore, even in the case of physical quantities, it is, today, customarily accepted that ordinal properties can be measurable (the VIM3 [International Vocabulary of Metrology] calls them "ordinal quantities" ».*

2.3.1 Que veut dire mesurer une propriété mentale ?

Si nous reprenons la définition donnée par Mari (2013) sur le cadre formel de la mesure « *To measure a quantity means to define a unit and to establish how many times the unit is contained in the given quantity. The measurement result is expressed by a number, which expresses how many times the given quantity is greater (or possibly smaller) than the selected unit* », nous ne pouvons nous référer à ce cadre puisque nous ne pouvons avoir d'unité avec les propriétés mentales ni vérifier empiriquement les axiomes nécessaires d'une mesure.

2.3.1.1 Une représentation

Mais ne pas pouvoir se référer au cadre formel de la mesure n'interdit pas de poser une mesure comme représentant une propriété existante pour une propriété mentale. En effet, le cœur de la discussion est sur ce qu'est une mesure. Mari (2013) « *Even more fundamental, because operating at a philosophical level, is the opposition which can be transversally found in these definitions: is measurement a determination, i.e., the discovery of an entity existing independently of the measurement, or an assignment, i.e., the choice of an entity according to information and criteria derived from the measurement itself?* » (p. 2891). En accord avec bien des auteurs, et donc en opposition avec Michell par exemple, je considère qu'une mesure n'est pas la découverte, mais la simple représentation d'une propriété. Ce positionnement découle logiquement de tout ce dont j'ai discuté sur la réalité et notre manière de l'appréhender. Je considère donc que la RTM est le cadre formel de référence nécessaire en psychologie. La RTM pose qu'une mesure n'est pas la découverte d'une vraie valeur réelle d'une propriété, mais une procédure d'assignement d'une valeur en correspondance avec notre connaissance de la propriété considérée. Le cadre formel à utiliser dès lors est un cadre

permettant de valider un isomorphisme (ou homomorphisme) entre une propriété et des valeurs numériques, non pour prétendre mesurer LA vraie valeur d'une propriété, mais pour représenter cette propriété par une valeur de façon fiable au sens du respect des axiomes d'une mesure.

2.3.1.2 Le problème de la protéiformisation

Des auteurs, Ziv & Hotam (2015) par exemple, posent le problème fondamental d'un tel isomorphisme posé par la RTM car pour eux les propriétés mentales sont protéiformes, donc en contradiction avec une mesure qui les réduit à une réalité unique. Leur critique de la mesure est que la littérature considère la mesure est comme le « gold standard » et les résultats empiriques deviennent « la vérité » (« Truth »). « *As researchers, if we abandon the single measure/gold standard approach in favor of a multilayered assessment approach, we can better broaden our fields.* » (p. 287) et « *because human psychology is so inherently complicated and multifaceted, it is unlikely that any measure is capable of capturing a large percentage of the psychological state it aims to assess [...] Such critical awareness invites us to remember that our overreliance on measures is unavoidably informed by the contemporary social and historical conditions we live in, and which induced the distancing of the field from any former faith that was invested in theory and in theorizing.* » (p. 287). Pour autant pour ces auteurs, la solution n'est pas de plus mesurer, mais avoir une adaptation rigoureuse de la mesure à la complexité des propriétés mentales: « *the solution may not be in the abandonment of the quantitative approach in favor of a qualitative approach, but rather in adapting a more rigorous quantitative or multi-method approach that acknowledges and respects its limits vis-à-vis the complexity and multifaceted nature of the psychological phenomenon* ». Il faut donc pour eux poser une mesure d'un objet protéiforme.

2.3.1.3 L'objectivité et vérité

De fait, les propriétés mentales apparaissent protéiformes au regard des conceptualisations rarement homogènes de la littérature. Si nous discutons en prenant comme cadre l'empirisme-réaliste, cette protéiformisation des propriétés mentales est en contradiction avec la nature même de la propriété mentale car cette dernière devrait être un objet externe et fixe, donc non-protéiforme. C'est à mon sens toute la contradiction de ce cadre épistémologique de l'empirisme-réaliste qui est incapable de poser des concepts définitifs des objets auxquels il se réfère. Mais je pars de l'approche proposée, le pragmatisme-empirique, en considérant les

propriétés mentales comme des objectivités « pour nous », dans le sens pratique et social. Ce « pour nous » est bien à comprendre au sens où Bourdieu donne au champ de la science, c'est-à-dire pour une communauté scientifique. Le pragmatisme se situe bien dans cette vision sociale : « *Or cette méthode, ce sont aussi les pragmatistes qui nous la livrent, une méthode qui prend en compte l'expérience et qui repose sur le modèle intersubjectif d'une communauté de chercheur [...] pour un pragmatiste, même la notion de "vérité" n'a aucun sens dans la "solitude morale"* » (Tiercelin, 2002).

Pour une communauté, une propriété mentale objectivée par ses manifestations et catégorisée dans un concept unique peut entraîner une procédure de mesure potentielle... autant que l'on considère cette procédure comme nécessairement renégociable dans le sens donné par Putnam. Cela n'est pas contradictoire avec une vision protéiforme d'une propriété mentale dans le sens où la négociation de cette propriété peut être différente pour une autre communauté, selon des approches différentes, des contextes ou enjeux particuliers. La mesure n'est pas une formalisation d'une vérité absolue, mais la formalisation d'une propriété objectivée comme unidimensionnelle pratique par une communauté. Cette mesure ne sera pas une « vérité » et ne pourra prétendre poser un cadre de mesure replicable sans questionnement. Ce sera juste une objectivation pratique « pour nous » d'une catégorie posée pour une communauté donnée.

Mais de nouveau, ceci ne tend pas vers le constructivisme. La contradiction entre différentes approches d'une propriété mentale doit être discutée relativement au cadre actualisé ; sa protéiformisation doit être justifiée. Or, le nœud le plus fréquent n'est pas tant une réelle protéiformisation d'une propriété mentale, mais la contradiction entre des approches « passées » et des approches « actualisées » d'une propriété mentale. C'est bien plus l'affirmation d'avoir posé un jour LA mesure d'une propriété qui rentre en contradiction avec l'évolution de son appréhension concrète ; et les formes différentes que prennent certaines propriétés mentales ne sont le plus souvent que le reflet de cette contradiction entre la réalité pratique et une vision empiriste-réaliste qui laisse croire que d'anciennes formalisations sont toujours valides. C'est bien plus un effet d'inertie et de conformisme qui génère cette protéiformisation qu'une réelle appréhension protéiforme d'une propriété à un instant *t*. « *Si nos concepts peuvent informer nos perceptions, ils peuvent aussi les contaminer* » (Tiercelin, 2002).

Poser une mesure comme étant une représentation d'une propriété mentale inscrite dans une pratique semble ainsi possible... pour autant que l'on ne considère pas de telles mesures comme « vraies » au sens généralisables en tout temps et tous lieux. Dès lors le recours à la

théorie représentationniste de la mesure ne semble pas contradictoire avec les propriétés mentales.

2.3.2 Retour à la RTM

La théorie *Representational Measurement Theory* (RTM) part d'une « représentation » numérique de la réalité, donc en rupture avec l'adhésion d'une réalité par nature numérique (Kyngdon, 2008b).

2.3.2.1 Le besoin d'un isomorphisme

La RTM considère qu'il y a un système relationnel empirique entre l'attribut et les valeurs. Comme le dit Sherry (2011) : « *Representation differs from description* », donc nous sommes bien dans un processus de représentation formalisé. Pour avoir une telle mesure, il faut un homomorphisme entre les qualités de l'attribut et une échelle numérique. McGrath (2005b) cite Krantz, Luce, Suppes, & Tversky (1971) : « *Most of the psychometric literature is based on numerical rather than qualitative relations (e.g. matrices of correlation coefficients, test profiles, choice probabilities) [...] For example, in scaling aptitude, intelligence, or social attitudes, test scores or numerical ratings are usually interpreted as measures of the attribute in question. But in the absence of a well-defined homomorphism between an empirical and a numerical relational structure, it is far from clear how to interpret such numbers* ». (pp. 32–33). C'est à mon sens en effet dans cette définition claire d'un homomorphisme que se situe le vrai problème du mesurage.

2.3.2.2 Même en physique...

Il faut impérativement vérifier cet homomorphisme entre la propriété mentale et une échelle numérique. Les mesures conjointes ne sont qu'un cadre formel de validation d'un tel homomorphisme potentiel entre un phénomène et des valeurs numériques. Le cadre des mesures conjointes n'est guère applicable en psychologie pour les raisons évoquées précédemment (les erreurs de mesures entre autres). Il n'y a donc pas de moyens formels de valider un tel homomorphisme. Mais Markus & Borsboom (2013a) considèrent que la démarche empirico-pratique en physique butte souvent sur le même problème que la psychologie. Même pour la physique, le cadre formel de la mesure est une idéalisation. En effet, Sherry (2011) renvoie la critique du cadre formel de la mesure adressée aux psychologues (comme Michell le fait) aux problèmes pratiques du cadre formel en physique :

« *Structural similarity in the sense of mathematical isomorphism is too strong a requirement, and the situation in contemporary physics supports this claim [...] Measurement theory analyzes procedures used to derive quantitative measures from qualitative observations [...] a representation theorem requires us to treat a collection of concrete observable things as a set in the strict mathematical sense [...] In truth, a representation theorem describes the idealized assumptions we bring to a genuine empirical relational system in order to produce a hypothetico-deductive system that can take advantage of the inferential powers of arithmetic* »

Car, de fait, même les propriétés physiques considérées comme quantitatives ne vérifient pas les axiomes formels « *length and distance are approximately continuous* » (Sherry, 2011). Falissard (2008b) souligne que « *les physiciens des particules reconnaissent souvent qu'il leur est impossible de définir véritablement ce qu'est une mesure, tout en précisant, cependant, que cela ne les empêche nullement d'en réaliser* » (p. 7). De même pour l'analyse conjointe, Sherry (2011): « *Plainly a conjoint system depends upon idealization in exactly the ways an extensive measurement does. Here too, our belief that a particular domain approximately satisfies a conjoint system does not rest upon an inductive generalization from observations of the domain. It rests, instead, upon an objective inference from the benefits of treating the objects as though they satisfy the axioms* ».

2.3.2.3 Une démarche pragmatique

Donc, la démarche pragmatique est pour Sherry (2011) au cœur de la mesure même en physique, basée sur une abduction, c'est-à-dire des hypothèses émises de mesurabilité, ensuite passée au crible de la démarche déductive et inductive, avec un effet de boucle critique, se référant à une assertibilité garantie : « *both the claim that temperature is a continuous quantity and the claim that it is measured by Fahrenheit's mercury in glass thermometer rest on abductive grounds [...] our justification for applying real numbers to the domain is abductive* » (p. 517). Vouloir imposer un cadre formel renvoie à l'impossibilité pratique de valider n'importe quelle mesure, « *The history of temperature suggests a middle path between embracing a research program dictated by measurement theory and simply abandoning the scientific task of quantification.* » ; et Sherry encourage les psychologues à adopter une démarche pragmatique (comme celle de Black pour la température) : « *It is open to psychologists to be encouraged or discouraged in their attempts to quantify mental attributes by comparing their accomplishments with Black's* ». L'idée n'est pas d'accepter une quantification sans discussion, la mesure doit s'inscrire dans une démarche pragmatique, donc

objectivée, telle que Sherry (2011) le pose : « *Acceptance of the pragmatic approach to quantification does not, of course, entail the existence of quantitative psychological attributes. The pragmatic value must be demonstrated.* » (p. 253).

La démarche reconnue pragmatique liée aux propriétés mentales permet de contourner le problème de la recherche de la « vérité » quantitative sous-jacente à la mesure d'une propriété mentale car une telle « vérité » n'est pas l'objet de notre démarche. Nous devons questionner, relativement à notre connaissance et pratique du phénomène étudié, si celui-ci est mesurable ; c'est-à-dire s'il répond à une propriété à même d'être représentée par une mesure, et si oui laquelle. En psychologie, la mesure est justifiée par son assertibilité garantie, donc en permanence questionnable car basée sur le cadre de nos connaissances et de nos pratiques. L'opérationnalisation de la mesure associée aux propriétés mentales est en phase avec la conceptualisation : c'est une démarche pragmatique-réaliste de représentation d'une propriété.

2.3.3 Que veut dire attribuer une valeur numérique ?

La discussion centrale sur la mesure est la critique de l'affectation de valeurs quantitatives à des propriétés mentales. Bien des auteurs considèrent qu'il faudrait user essentiellement de mesures ordinales ou nominales comme je l'ai déjà dit (Sijstma, Michell par exemple).

2.3.3.1 La densité

La plupart des auteurs s'accordent à considérer que la plupart des propriétés mentales sont au minimum des propriétés ordinales (on peut leur attribuer une propriété de « plus » ou « moins » sans pour autant pouvoir attribuer une échelle numérique). Ces auteurs, avec raison, pointent le fait que d'une variable ordinale on ne peut pas passer sans discussion à une variable numérique comme le font la plupart des auteurs usant de mesures en psychologie. Michell (2012) rappelle que cette validation ne peut se faire uniquement en considérant que la propriété est « plus » ou « moins » présente chez une personne car cela n'implique qu'une relation ordinale et non numérique. Il faut justifier un homomorphisme entre la propriété et une échelle numérique. Donc, il faut discuter de la capacité à représenter la propriété mentale sur un continuum numérique. Le recours au cadre formel de la mesure (ou mesures conjointes) n'est pas possible en psychologie et donc interdit cette vérification. Mais je considère, en phase avec Sherry (2011), que concrètement la démarche doit être autre car on ne peut jamais prouver empiriquement que des données vérifient le cadre strict de la logique-mathématique (qui demande une infinité de résolutions).

La différence entre une variable ordinale et une variable continue peut être comprise comme un passage d'une variable ordonnée, respectant donc un ordre simple, à une variable respectant une densité au sens mathématique²⁵. Luce & Suppes (2002) considèrent en effet que « *La mesure en plus d'une relation d'ordre peut se définir par la densité d'un objet. La propriété archimédienne assure une forme de densité* ». La question de la légitimité de considérer une propriété mentale comme représentée sur une échelle numérique envoie à la légitimité de considérer une propriété mentale comme dense, que l'on peut considérer comme pouvant être représentée sur un continuum numérique, c'est-à-dire vérifiant une « continuité » sur un intervalle. Or Sherry (2011) Remarque : « *The Archimedean condition is likewise impossible to verify by simply observing the behavior of the domain* ».

2.3.3.2 La continuité empirique

Partant d'une démarche pragmatique, Sherry (2011) considère qu'il faut montrer qu'une propriété est "approximativement quantitative" par le fait qu'elle est "empiriquement continue" : « *Satisfaction of something like the conditions of empirical continuity in Körner (1962) will suffice for measurement [...] When Michell observes that predictive failure of weaker tests won't rule out the hypothesis that an attribute is quantitative, he implies that the stronger tests made possible by measurement theory could rule out the hypothesis that an attribute is quantitative. But it's a myth to think such a test might once and for all rule out an hypothesis of quantity [...] If treating an underlying attribute, say intelligence, as quantitative allows one to employ mathematics to bring order to the data, then a plausible explanation for this success is that the attribute is approximately quantitative.* »

Körner (1962) différencie en effet le concept mathématique et le concept empirique de « continuité ». En mathématique, une classe est divisible en une infinité de sous-classes, ce qui est impossible empiriquement ; il est donc impossible empiriquement de montrer la densité d'une propriété. Le modèle mathématique est une idéalisation. Pour Körner, un ensemble empirique est dense s'il est divisible entre les différents éléments, c'est-à-dire pour Körner (1962) « *the notion of an empirically continuous series of classes P1, P2, ... Pn serves equally to explain the notion of continuous qualitative change of one thing and of a continuous gradation of a number of things* » (p. 13). C'est une démarche "subjective" pour Körner, prenant par exemple le spectre des couleurs qui peut être appréhendé comme un

²⁵ Je ne développe pas la densité dans sa formalisation mathématique car cela renvoie à des notions de topologie qui demanderaient des développements longs. Mais intuitivement ce terme de « densité » se comprend sans le recours au formalisme mathématique.

spectre continu ou comme une variable ordinale catégorisée selon les personnes. La continuité empirique renvoie donc à une densité empirique au sens d'une objectivation rationnelle d'une graduation continue par une démarche rationnelle empirique... mais bien inscrite dans une pratique sociale car « subjective ».

2.3.3.3 Valider la continuité empirique

Valider empiriquement un continuum fut l'objet des approches psychométriques à leur émergence (voir O. Martin, 1997). Ma position ne se prétend pas dans la continuité de ces travaux (Spearman, Thurstone, ...) dans le sens où je me positionne dans une démarche pragmatique sans prétendre mesurer un objet « vrai » qui serait en tout temps et en tout lieu mesuré de la même manière. De ce point de vue, ce que Hacking développe dans « *l'âme réécrite* » (Hacking, 2006) sur la mesure et les moyens de la valider me paraît plus proche de mon approche.

Hacking pose qu'il faut vérifier l'hypothèse *d'absence d'écart* c'est-à-dire concrètement que sur tout segment de l'échelle, je peux théoriquement trouver une personne vérifiant cette valeur. La démarche subjective de la continuité empirique revient à cela. Mais on peut user de tests statistiques pour vérifier si une distribution empirique peut être considérée comme non-statistiquement différente d'une distribution théorique. La difficulté est de poser la loi théorique. Usuellement on utilise la loi normale, mais comme le dit O. Martin (1997), la loi normale est passée d'un « *outil pratique* » à « *un statut ontologique* ». Il faut donc justifier ce choix. Les tests de Kolmogorov-Smirnov ou de de Shapiro-Wilk peuvent par exemple être utilisés car ils testent l'adéquation d'une distribution empirique et d'une distribution théorique (tests non-paramétriques, donc utilisables avec de petits échantillons). Mais un test *visuel* à partir de la distribution empirique peut suffire. L'objectif est de vérifier empiriquement qu'il n'y a pas de segments sans valeur.

Hacking pose qu'il faut une *absence de seuil*, c'est-à-dire avoir une échelle où le score nul n'existe pas afin de ne pas avoir deux catégories : ceux qui ont « zéro » au test et les autres. L'hypothèse sous-jacente est que le test mesure une quantité que tout le monde a plus ou moins. La tension est dès lors sur le nombre de questions d'un test car plus ce nombre sera faible, plus la probabilité sera élevée d'avoir des personnes ayant un score de zéro. Il faut donc vérifier si un test ne discrimine pas un sous-groupe au seuil de zéro.

Enfin, Hacking pose l'hypothèse de la *régularité*. De la manière dont hacking pose cette hypothèse, on devine un renvoi à la loi normale qui, comme je viens de le dire, a aujourd'hui

un statut ontologique. Sans se référer au cadre gaussien, on peut vérifier visuellement si la distribution est multimodale, avec des groupes séparés homogènes, ou bien user de procédures typologiques (CAH, K-means, ...) afin de vérifier qu'il n'y a pas des sous-groupes distincts ayant des structures très homogènes entre eux et très hétérogènes entre les groupes (variabilité intra vs inter).

Enfin, une mesure ne doit pas être sensible à l'instrument, cela veut dire concrètement que l'on doit obtenir le même résultat avec deux instruments différents. Cela renvoie à *l'invariance* de la mesure (« *for a scale to be considered to be valid, the values assigned to the items must not depend on the opinions of the persons used to calibrate the scale* », définition de Leplège, 2003a, p. 457). Or il y a toujours un risque en psychométrie de réifier une valeur numérique obtenue sans réels fondements, c'est-à-dire attribuer artificiellement une valeur numérique. Pour remédier à ce problème, comme toutes les mesures sont effectuées à partir d'items sélectionnés, il est a priori possible de construire différents instruments d'une même propriété mentale pour valider sa robustesse. Cela revient à l'approche multitraits-multiméthodes (MTMM), qui peut être employée concrètement et qui pourtant pourrait être une procédure utile.

Ces vérifications empiriques sont certainement nécessaires à effectuer pour valider un continuum hypothésé a priori.

2.3.4 Comment attribuer une valeur numérique ?

Comme je l'ai dit, c'est un isomorphisme qu'il faut caractériser entre la propriété mentale et l'échelle numérique qui la représente. La démarche pour valider cet isomorphisme entre une propriété et une échelle numérique est elle aussi une démarche pratique.

2.3.4.1 Convergence et divergence avec la littérature académique

Posé ainsi, je suis, peut-être de façon surprenante, à priori en phase avec la littérature académique. Les auteurs centraux dans la discussion (Sijtsma, Michell, Trendler, Kyngdon par exemple) discutent en effet de l'homogénéité vs l'hétérogénéité des indicateurs d'une propriété mentale. Pour eux si les indicateurs d'une propriété sont homogènes et comparables quantitativement, nous pouvons appliquer une représentation numérique de la propriété mentale ; sinon, si les indicateurs diffèrent qualitativement et non quantitativement, il faudra avoir recours à une mesure ordinale voire nominale (Kyngdon, 2013; Michell, 2012). Donc,

comme le dit Sijtsma (2012b), c'est la « théorie » qui pilote, ce qui revient à avoir recours à une démarche subjective, inscrite dans la pratique concrète de la communauté à laquelle se réfère la théorie, pour poser une représentation de la propriété mentale par une mesure continue sur un intervalle. Michell (2012) illustre cela par le test de la *dépendance fonctionnelle* des personnes âgées dont les items (degrés de gravité) sont : (1) *climbing stairs*, (2) *transferring to bathtub*, (3) *bathing*, (4) *walking*, (5) *dressing upper body*, (6) *independent toileting*, (7) *transferring to bed*, (8) *dressing lower body*, (9) *mobility without a wheelchair*, (10) *bladder control*, (11) *performing personal grooming*, and (12) *bowel control*. Le fait que des données avec le modèle de Rasch soient très fiables pour ces indicateurs ne veut pas dire que nous avons plus qu'un ordre ici. En effet, si au sein d'un item (comme « *climbing stairs* »), les variations sont homogènes (donc quantifiables), le passage d'un degré à l'autre (comme le passage de *climbing stairs* à *transferring to bathtub*) est qualitatif (hétérogène) car on ne peut définir des comparaisons quantitatives entre des différences de degrés pour ce test pour Michell.

Ce qui est intéressant à mon avis, c'est que Michell (2012) considère sans le discuter réellement que chaque degré d'un test puisse être certainement représenté par un attribut quantitatif, ce n'est que le passage d'un degré à l'autre qui est qualitatif car nous ne pouvons pas « mesurer » les écarts d'un degré à l'autre. Michell caractérise ainsi la possibilité de pouvoir « considérer quantifiable » un attribut à partir d'un raisonnement rationnel sans pouvoir le démontrer. De même pour Trendler (2009) des manifestations peuvent révéler des différences hétérogènes non mesurables, d'autres au contraire peuvent être identifiées comme certainement homogènes et quantifiables. Kyngdon (2013) illustre des tests dont les items sont homogènes, c'est-à-dire que la différence entre chaque item peut être mesuré de façon fiable (modèle d'utilité de Tversky et Kahneman : *Lexile Framework for Reading*). Kyngdon (2013) discute de la « difficulté » de certains tests usuellement modélisés par le modèle de Rasch, et montre que la difficulté ne peut être utilisée comme une échelle d'intervalle pour des tests d'habileté cognitive comme *Trends in International Mathematics and Science Study* (p. 238-240) ou la mesure d'attitude (p. 240-243). Pour Sijtsma & Emons (2013) le recours à des mesures ordinales ou nominales (classes) est certainement à recommander « *latent class model analysis showed that proportional reasoning is best represented by a limited number of ordered latent classes (Jansen & Van der Maas, 2001) and transitive reasoning is best measured using a non-parametric, ordinal IRT model rather than a more restrictive parametric IRT model (Bouwmeester & Sijtsma, 2004; also see Michell, 2012, on ordinal attributes).* » (p. 788).

Mon objectif n'est pas de discuter de mesures particulières, donc je ne cherche pas à vérifier si pour telle ou telle mesure ces auteurs ont raison de poser une mesure ordinale. Je cherche à montrer que la démarche proposée pour ces auteurs, proposant un cadre objectif de la mesure en psychologie, correspond à une démarche pragmatique ; c'est-à-dire rationnelle et empirique, inscrite dans un cadre pratique (le cadre théorique entre autres), donc subjective, mais sans pour autant s'opposer à une forme d'objectivité.

Ce que je pose n'est donc pas en contradiction avec la littérature académique dans la pratique concrète, mais fondamentalement en contradiction avec les soubassements épistémologiques. D'une part la plupart de ces auteurs, et bien d'autres, questionnent les processus en amont pour savoir si une propriété mentale est mesurable en s'appuyant sur l'homogénéité ou hétérogénéité des processus générant la propriété mentale. Pour eux, si les processus sont non homogènes, la propriété ne sera pas mesurable. C'est la confusion déjà énoncée entre les processus cognitifs et la propriété émergente. De plus, pour la plupart de ces auteurs la validation semble à priori déconnectée de la pratique concrète car ils se réfèrent à une propriété théoriquement existante indépendamment de leur pratique... tout en se remettant à une démarche très pragmatique de validation.

2.3.4.2 La démarche pragmatique proposée

Je considère qu'à l'image de ces pratiques concrètes discutées dans la littérature, il faut questionner une propriété mentale pour savoir si elle représente un domaine dense au-delà d'un simple ordre. Cette procédure ne peut se valider par un recours à un formalisme (Hölder ou Mesures Conjointes) mais par une discussion : la propriété telle que catégorisée par le cadre théorique (du moment) est-elle une propriété qui ne peut se décliner en « plus » ou « moins » présente ? Si oui nous pouvons poser au minimum une mesure ordinale, sinon elle sera nominale (il n'y a pas de comparaisons possibles des différents états de cette propriété).

Si nous avons un ordre, celui-ci peut-il être considéré avec de potentielles ruptures qualitatives, c'est-à-dire une hétérogénéité entre les niveaux qui les rendent incomparables quantitativement ? Si oui, il faudra utiliser une mesure ordinale sans recourir à une mesure quantitative. S'il n'y a pas hétérogénéité entre les états de la propriété mentale, peut-on considérer qu'il y a potentiellement un continuum de la propriété représentée sur une échelle numérique ? Cela revient à vérifier subjectivement une continuité empirique de la propriété mentale (une densité empirique) au sens de Körner (1962). Si cela est pratiquement possible, nous pouvons poser une représentation numérique entre la propriété mentale et un intervalle

numérique et considérer un homomorphisme entre les deux. Définie ainsi, une mesure vérifie les conditions de la RTM : homomorphisme entre une propriété et une échelle numérique. L'unicité de cette échelle est vraie à une transformation près (l'unité étant arbitraire, on peut générer une infinité d'isomorphisme mais équivalents deux à deux par simple transformation linéaire). Une telle échelle représentera « fidèlement » (Luce & Suppes, 2002) la propriété mentale au sens où les valeurs numériques auront un sens relativement au cadre pratique de l'objectivation des manifestations.

Mais une validation a posteriori par des tests (statistiques ou visuels, comme je l'ai exposé précédemment) est certainement nécessaire.

2.3.4.3 Critique de la quantification non-critique

Michell, Sijtsma, Trendler, Kyndgon et d'autres ont raison de pointer que cette discussion sur la nature quantitative d'une propriété mentale est rare dans la littérature académique utilisant des mesures de propriétés mentales. La plupart des travaux usent de mesures quantitatives sans questionner son objectivité. Pour autant, une quantification me paraît possible dans une pratique concrète, c'est-à-dire par une approche pragmatique-réaliste ... autant que l'on considère une telle mesure comme renégociable.

Vautier, dans différents articles, critique, avec raison, la généralisation de mesures quantitatives sans questionner leurs fondements, pour exemple, Vautier (2015a) : « *Pour autant qu'on sache décrire les manifestations de grandeurs comme l'intelligence ou l'anxiété dans un certain langage d'observation psychotechnique, ce langage n'est pas le langage du "plus ou moins", mais celui des m-uplets. La plupart des variations qu'on sait envisager de façon empirique ne se laissent appréhender ni en "plus", ni en "moins". L'erreur scientifique des psychologues qui croient aux grandeurs comme l'intelligence ou l'anxiété est qu'ils n'admettent pas que les phénomènes qu'ils savent étudier sont d'abord des phénomènes qualitatifs (les réponses aux items des tests). Au lieu de respecter la forme de manifestation de ces phénomènes, ils les déguisent en nombres (ou scores), vendant ainsi la peau de l'ours avant de l'avoir tué : ils font comme si on pouvait parler scientifiquement en "plus ou moins" de l'intelligence ou de l'anxiété de n'importe qui, sans que cela pose le moindre problème d'authenticité [...] Mais la réalité, telle qu'elle se manifeste dans nos cadres descriptifs, nous apprend-elle que le langage du "plus ou moins" est approprié pour parler authentiquement des manifestations de l'intelligence ou de l'anxiété de n'importe qui ?* ». Je considère, j'y reviendrai dans la discussion sur les probabilités, que l'on ne doit pas s'interdire de passer de

m-uplets à une vision quantitative potentielle. Mais de fait, en accord avec Vautier, ce glissement de m-uplet à une quantification s'effectue le plus souvent sans questionner ce glissement.

Tout n'est pas mesurable en psychologie au sens de pouvoir attribuer une valeur numérique. Mais de cette critique fondamentale de la psychométrie telle que pratiquée ne renvoie pas à mon avis à une interdiction de poser une démarche de mesure quantitative potentielle, à justifier et sans la considérer comme « vraie », mais comme « objective » relativement au contexte et enjeux. Vautier (2015g) considère d'ailleurs « *cette méthodologie est néanmoins utile pour répondre à certaines questions de débroussaillage* ».

2.3.5 Validité des mesures subjectives ?

La plupart des mesures en psychologie s'effectuent par des procédures auto-administrées, qui demandent aux personnes de s'autoévaluer. Une question soulevée par les mesures en psychologie est donc le recours le plus souvent à la subjectivité de la personne.

2.3.5.1 Les problèmes des mesures subjectives

L'introspection est souvent critiquée, avec toute une littérature pointant les problèmes divers soulevés (voir par exemple Petitmengin, 2009; Silvia & Guido, 2001). Pour Silvia & Guido (2001) la question se pose de savoir si les personnes sont capables d'auto-jugement, cela renvoie à la capacité de la personne d'avoir conscience d'elle-même, de ses pratiques. Le problème posé ainsi est la différence entre une propriété « objective » et une vision « subjective » de la personne de sa propriété. Petitmengin (2009) rappelle que ce problème de l'introspection est ancien, comme le notait déjà A. Comte : « *L'individu pensant ne saurait se partager en deux, dont l'un raisonnerait, tandis que l'autre regarderait raisonner. L'organe observé et l'organe observateur étant, dans ce cas, identique, comment l'observation pourrait-elle avoir lieu ? Cette prétendue méthode psychologique est donc radicalement nulle dans son principe* » (Comte, 1830). En conséquence, « *Many psychologists are skeptical of the value of self-reports* » (Silvia & Guido, 2001, p. 23).

Le problème central posé est qu'une personne ne s'observe pas comme un objet, elle s'observe dans un rapport dialectique avec elle-même et dans un objectif (plus ou moins conscient), avec des distorsions potentielles : le problème du langage qui ne peut concrétiser de façon simple des états mentaux ; le problème de la réalité vs l'invention d'une réalité pour la personne ; le recours à une métaconscience difficile à établir ; l'impossibilité de vérifier si

l'exposé est vrai ; le fait que l'introspection modifie la personne, ... Petitmengin (2009) cite Schlanger : « *As introspection is not a rigorous method, one must not expect any scientific results from it. By using it one cannot hope to reach results of observation and experimentation that would be repeatable and controllable, in the way public observations and experiments in physics or in chemistry are controllable - since mental phenomena that introspection observes are private, inner, non public and communicable only by the means of language, by which one expresses them* » (p. 372).

2.3.5.2 Ce que révèle la mesure subjective

Je ne peux qu'adhérer à toutes ces limites de l'introspection... mais comme le considère Schlanger ces critiques sont relatives à une démarche d'objectivité par rapport à un objet existant "en soi" comme en physique. Il faut à mon avis repositionner l'introspection dans la démarche globale de la connaissance en psychologie. L'introspection n'est pas la révélation d'un objet mental fixe mais plutôt « *inasmuch as they contain information about how the self relates to the environment* » (Silvia & Guido, 2001, p. 23). Petitmengin (2009) rappelle que l'introspection est une procédure de description de notre expérience mentale. Voir l'introspection comme une capacité de se décrire précisément est une vision « naïve » de l'introspection (Petitmengin, 2009). Il faut donc non pas chercher le « vrai » mais « l'authenticité » (Petitmengin, 2009). L'introspection ne révèle pas « la réalité » mais « l'expérience vécue » de la personne. Le langage est le moyen de se conscientiser, et les personnes sont des êtres conscients d'eux-mêmes, avec toutes les distorsions de la conscience de soi. L'introspection est donc à comprendre dans l'interaction sociale comme procédure subjective de s'objectiver. L'introspection révèle la pratique sociale de la personne en tant que sujet, certes enfermée dans un cadre social linguistique et de catégories réifiées. Toute procédure d'introspection (questionnaire par exemple) n'est donc pas un protocole de mesure d'un objet, mais un dialogue, à comprendre dans ce cadre de relation sociale. « *The validity of a first person report is a validity 'in action', which cannot be measured in static terms of correspondence between the report and the experience, but in dynamic terms of performative consistency of the acts which produce it* » (Petitmengin, 2009, p. 400).

2.3.5.3 L'intérêt des mesures subjectives

Falissard (2008b) revient longuement sur cette question centrale de la mesure subjective « *ce qui nous intéresse, c'est bel et bien ce que ressent le sujet et non les scores obtenus par des*

instruments dont on ne peut interpréter le sens intrinsèque [...] un attribut subjectif ne correspond à rien de concret, tout juste à un mot, qui plus est au sens généralement imprécis [...] Il serait donc illusoire d'espérer en obtenir une mesure scientifiquement acceptable. Prenons l'exemple de la tristesse. Il est vrai que, curieusement, il est difficile d'en obtenir une définition, mais est-ce pour cela que le sens de ce mot est imprécis ? Bien au contraire. La tristesse est un sentiment élémentaire que tout être humain a déjà éprouvé. Il suffit, en fait, que je dise : "je suis triste..." pour que n'importe quelle personne sache exactement ce que je ressens. [...] par définition, un élément subjectif est relatif à ce que ressent un individu ; or, il est impossible d'avoir la moindre preuve qu'une personne ressent bel et bien ce qu'elle dit ressentir (si tant est que "ressentir bel et bien quelque chose" ait un sens). Il serait donc de nouveau impossible d'espérer obtenir une mesure scientifiquement acceptable de l'élément subjectif en question. Trois objections peuvent néanmoins être formulées à l'encontre de cet argument [...] Pour des questions sans véritable enjeu, si la réponse est recueillie dans un environnement de confiance, il est peu probable que le sujet exprime quelque chose de véritablement différent de ce qu'il éprouve [...] Il est vrai que le niveau actuel de précision des mesures subjectives (en santé tout particulièrement), quand on peut le déterminer, est souvent médiocre ou mauvais. Mais cela est-il rédhibitoire quant à la scientificité des disciplines de la subjectivité ? Rien n'est moins sûr : d'une part certaines mesures objectives sont d'un niveau de précision équivalent (la tension artérielle mesurée au brassard en est un exemple) ; d'autre part, les mesures subjectives sont susceptibles de progresser, et ce d'autant plus que l'on observe une évolution parallèle des paradigmes de mesure dans les disciplines de la subjectivité et dans les sciences les plus « objectives », la physique notamment » (p 4-8).

En conclusion, et en phase avec Falissard, il faut clairement parler de mesure « subjective », c'est-à-dire rompre avec la prétention de mesurer « la réalité » au sens biologique, mais prétendre juste mesurer le « vécu » de la personne au sens de sa conscientisation, c'est-à-dire mesurer une propriété mentale dans sa fonction sociale. Un protocole de mesure faisant appel à une auto-évaluation, s'il est construit de sorte à avoir une réelle procédure d'auto-évaluation (confiance et implication de la personne entre autres), et s'il est compris dans un rapport dynamique et non avec une vision statique de la personne, permet d'objectiver la personne en tant que sujet social. Cependant, comme le souligne Falissard, le recours à des mesures hétéro-administrée (« terme signifiant que la mesure sera effectuée à partir de l'observation d'un tiers et non à partir de l'évaluation du sujet par lui-même », définition de Falissard). peut-être une approche efficiente dans le contexte de troubles psychiatriques, en comprenant

de même ce protocole comme inscrit dans un cadre social (les catégories référentes des experts évaluant la personne) ; « *Le principe d'une mesure subjective obtenue par hétéro-évaluation repose sur une aptitude humaine largement partagée : l'empathie [...] en psychiatrie c'est souvent une évidence (un patient mélancolique et délirant aura bien du mal à se représenter et à rapporter avec justesse la qualité de ses émotions) ; plus généralement, le fait même d'être malade est susceptible de biaiser l'appréciation que le sujet porte sur certains de ses états propres subjectifs.* » (p. 10). Ce sont bien des indicateurs sociaux que nous utilisons ainsi pour mesurer une propriété mentale, donc orienté vers sa fonction sociale.

2.3.5.4 L'homogénéité du référentiel

Cette approche de la mesure subjective renvoie au sens des questions d'un questionnaire. Engelhard (2008) par exemple considère que des items doivent avoir un sens invariant pour les répondants, s'appuyant sur Guttman « *One of the fundamental problems facing research workers . . . is to determine if the questions asked on a given issue have a single meaning for the respondents. Obviously, if a question means different things to different respondents, then there is no way that the respondents can be ranked . . . Questions may appear to express a single thought and yet not provide the same kind of stimulus to different people.* » (p. 161). Or, de fait, une question posée ne peut avoir un sens strictement uniforme car elle est interprétée par la personne. La seule chose que l'on puisse demander à une question est d'être inscrite dans un référent homogène pour les différentes personnes, et permettre de représenter l'expression individuelle dans ce référent ; c'est-à-dire dans une catégorisation donnée des propriétés mentales. Les questions doivent donc s'inscrire dans une théorie et pratique socialement admise dans le sens où le questionnant et le répondant sont inscrits dans un dialogue qui nécessite une homogénéité de référence.

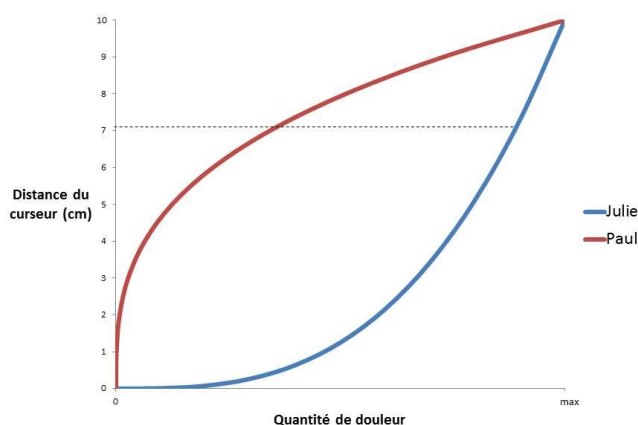
2.3.6 L'exemple de la douleur

Je conclus cette partie par un exemple repris de Vautier (Vautier, 2015e). Vautier (2015e) illustre, par la mesure de la douleur, l'impossibilité de pouvoir prétendre mesurer objectivement une propriété mentale. Cet exemple permet de caractériser ce qui est au cœur de la discussion que je mène contre cette critique.

2.3.6.1 L'exemple développé par Vautier (2015)

Vautier (2015e) pose ce graphique comparant Julie et Paul avec en abscisse « la quantité de douleur » et en ordonnée « la réponse, en cm, du ressenti de la douleur ».

FIGURE 12 : L'EXEMPLE DE LA DOULEUR



Tiré de Vautier (2015e)

Vautier discute de la comparaison de la douleur entre les deux personnes et considère : « *Sauf quand leur douleur est nulle ou maximale, lorsque Paul et Julie indiquent l'intensité de leur douleur par une même mesure, la douleur de Julie est plus grande que la douleur de Paul (ce qui, en termes courants, signifie que Julie est plus résistante à la douleur que Paul). Ainsi, si on admet que les fonctions de mesurage sont propres aux personnes, on ne peut pas utiliser les mesures pour comparer les quantités mesurées chez les différentes personnes : chacun est son propre instrument de mesure et on ignore comment passer d'un instrument à un autre.* » ; et discutant de la douleur de Paul : « *En d'autres termes, la fonction de mesurage ne conserve pas l'unité de changement de la douleur, ou encore, la mesure n'est pas métrique mais seulement ordinale. Calculer une différence entre plusieurs mesures pour quantifier le changement de la douleur, c'est-à-dire connaître la quantité de changement de douleur, est une chose vaine.* ». Vautier considère qu'il faudrait une fonction affine afin que l'expression de la douleur en cm puisse représenter quantitativement la douleur réelle. Il conclut : « *la personne répond ce qu'elle veut parce que c'est une personne, au lieu de répondre de manière machinale comme un objet, ou encore, un système. Par conséquent, au lieu de chercher à fonder son interprétation de la réponse sur des arguments psychométriques ou*

métrologiques peu plausibles, le psychologue devrait plutôt considérer cette réponse comme un acte de langage, qui implique l'intentionnalité des interlocuteurs plongés dans un champ de significations possibles. Si, par exemple, la personne énonce "7 cm" hier et "6 cm" aujourd'hui, il est possible qu'elle dise "j'ai un peu moins mal aujourd'hui qu'hier" dans le langage du clinicien qui la fait s'exprimer par déplacements du curseur de l'échelle. Mais il est impossible de déduire de ces énoncés qu'elle a objectivement moins mal, si on n'accepte pas la théorie du mesurage qui est la prémisse nécessaire d'une telle déduction. »

2.3.6.2 Qu'est-ce que la douleur ?

Ce qui, à mon sens, pose problème dans le développement de Vautier c'est la considération qu'il y a une « douleur objective » et une « expression de la douleur » comme deux objets séparés. Je suis bien évidemment en accord avec Vautier pour considérer que l'expression de la douleur comme « *un acte de langage, qui implique l'intentionnalité des interlocuteurs plongés dans un champ de significations possibles* », ce que j'appelle « inscrite dans une pratique », c'est-à-dire compte tenu du contexte et des enjeux (au sens très large et pas que les enjeux immédiats, c'est-à-dire incluant le contexte social, historique, ... et les enjeux sociaux de la mesure effectuée). Je crois en effet que la douleur comme état mental n'existe pas indépendamment d'un contexte social. Le même évènement (tomber) peut provoquer un traumatisme (se casser une jambe) mais génèrera des états mentaux du ressenti de la douleur très différents selon des situations sociales : je suis seul et chute en escalade et je dois rejoindre le campement coûte que coûte, la douleur semble « s'oublier » (au regard des expériences concrètes vécues par des alpinistes) ; je suis renversé par une voiture, immédiatement entouré de témoins et de secours qui me réconfortent et accablent le chauffard, je peux libérer ma douleur car je n'ai aucune responsabilité ni risque (sauf si la peur vécue anesthésie la douleur). La douleur ne peut se caractériser simplement par un processus physicochimique indépendant de la situation sociale et du ressenti inhérent. On ne peut pas différencier une douleur « objective » d'une douleur « subjective » comme propriété mentale car la douleur comme propriété mentale est subjective, la douleur est un ressenti. Je renvoie par exemple aux propos racistes que Gould (1997) rapporte de scientifiques du XIX^{ème}, caractérisant les « noirs » comme insensibles à la douleur. C'est bien le contexte social qui laissait percevoir la réalité ainsi par les « blancs », et peut être la « résistance à la torture » par certains esclaves.

Je crois que le graphique proposé par Vautier est caractéristique de la démarche empirique-réaliste qui considère qu'il y a un objet réel au-delà de la conscientisation de la douleur, c'est-à-dire en quelque sorte un objet biologique indépendant de l'être social. Or cet objet n'existe pas si on discute de propriétés mentales. Cela n'a aucun sens de comparer par exemple des stimuli générant de la douleur et de mesurer le ressenti pour vérifier la validité d'une échelle de douleur biologique, car les mêmes stimuli peuvent générer des ressentis différents selon le contexte. Il n'y a donc pas d'abscisse, juste l'ordonnée, dans le graphique proposé par Vautier : le ressenti de la douleur ; et c'est ce que cherche à inférer une échelle analogique de douleur.

Vautier écrit dans un autre papier (relativement à la fatigue) : Vautier (2015d) « *L'important est que si on prétend mesurer la fatigue avec des réponses à un ou des items de fatigue, on soit lucide sur le fait que le vécu "fatigue" n'est pas pertinent. Et si on s'intéresse au vécu ou encore au ressenti plutôt qu'à une quantité, alors il faut se poser la question des conditions de possibilité d'une connaissance objective de ce ressenti — vaste programme, qui ne relève pas du mesurage.* ». C'est là où je diverge avec Vautier car on s'intéresse à un ressenti qui peut être représenté par une quantité si la propriété mentale peut être considérée pouvoir être représentée sur un continuum, et pour la fatigue cela me paraît opérationnalisable (dans le sens où un continuum numérique permet de représenter « fidèlement » un ressenti de fatigue sans devoir passer par une échelle ordinale). On aura une « connaissance objective du ressenti » si on considère l'objectivité d'un point de vue pragmatique, c'est-à-dire une « objectivité pour nous », contextualisé, et non comme une objectivité au sens empiriste-réaliste représentant un vrai objet « naturel » que nous pourrions approcher biologiquement indépendamment du ressenti de la personne.

La douleur ou la fatigue peuvent certainement être appréhendées biologiquement, indépendamment de l'être social. On peut certainement faire des relevés biologiques caractérisant « objectivement » un degré de « douleur », ou de « fatigue ». Mais la mesure alors sera différente de la réalité de la propriété mentale. On mesure alors l'être biologique et non l'être social. En tant que propriété mentale, c'est bien de la fonction sociale de la propriété mentale qu'il faut partir (comment l'individu se conscientise ou comme son environnement l'objective) et donc à partir des manifestations de ces propriétés au sens de leurs fonctions interactives pour l'être en tant que sujet social (dont le discours verbal peut être une forme).

2.3.6.3 L'objectivité de la douleur subjective

Avec l'exemple simple de la douleur, j'ai cherché à illustrer la différence entre l'empirisme-réaliste, à la quête d'objet « vrais »... et donc comme le dit Vautier ce qui emmène cette approche dans des contradictions inextricables pour mesurer ces objets « vrais », et le pragmatisme-empirique tel que je le considère, où la propriété mentale n'existe que par ses manifestations, et est appréhendée nécessairement par ces manifestations relativement à un contexte et des enjeux... donc en considérant que mesurer conserve une objectivité mais dans un sens social pratique.

Ziegler & Vautier (2014) considère « *For example, it is semantically absurd to say that we measure a psychological process: a process is not a quantity* », ce qui est vrai, et c'est ce que j'ai cherché à expliquer dans la discussion sur la nature ontologique d'une propriété mentale. L'empirisme-réaliste considère le plus souvent que la propriété mentale est ce processus complexe en amont de ses manifestations (donc comme si la propriété mentale avait une existence biologique indépendante de sa fonction sociale en tant que manifestée par la personne) ; et de ce point de vue Ziegler et Vautier ont raison : un processus ne se mesure pas... d'où les tentations (*Network Analysis*) d'oublier la propriété mentale car trop complexe à analyser en tant que processus complexe. De mon point de vue, dans la mesure d'une propriété mentale nous ne mesurons pas un processus mais nous mesurons la propriété émergente qui a une fonction sociale et qui se révèle par ses manifestations sociales.

D'une façon plus générale Vautier écrit (Vautier & Afzali, 2015) « *a respondent is not a system but a person: a main difference between a system and a person is that a person responds to the rating task according to her/his intention to put marks on the VASs, entailing that the psychologist should be aware of the fact that the assessment occasion is a conversational setting. A person cannot be a measurement instrument, since a measurement instrument - e.g., a thermometer - has no intentionality, no sense of behaving itself in a conversational setting.* ». Dans les différents travaux de Vautier nous retrouvons la formulation que l'expression de la mesure par une personne est une composition de fonction ($f \circ g$) où g est la vraie fonction de la propriété mentale et f la fonction de subjectivisation de cette fonction g . Ce formalisme est valide si on considère la propriété comme « réelle » au sens d'une réalité en dehors de l'interaction sociale. C'est dans cette critique de l'impossibilité de construire un instrument fiable d'une mesure d'un objet réel caché que je situe ma critique envers Vautier et ses co-auteurs. Je considère que la fonction sociale de la propriété mentale ne peut être considérée comme une fonction séparée de ses manifestations.

La catégorisation sociale d'une propriété mentale rend obligatoirement cette propriété mentale comme propriété conscientisée au sens de sa fonction sociale (soit par la personne, soit par son environnement). Si la propriété mentale n'a aucune manifestation, elle n'existe pas en tant que propriété mentale (en comprenant bien qu'un même comportement pourra être caractérisé différemment selon les contextes, donc que l'existence d'une propriété mentale dépend du contexte comme je l'ai expliqué précédemment).

C'est cette propriété dans sa fonction sociale que nous mesurons (via des manifestations), et non un objet en amont de cette expression de la propriété mentale. Il y a donc nécessairement une « intention » (au sens où la propriété mentale a une fonction sociale, consciente ou non) dans l'émergence d'une propriété mentale car la personne ne peut être considérée en dehors de ses interactions sociales (d'où mon adhésion à l'intentionnalité développée par Searle pour les objets sociaux). Pour reprendre les exemples de la douleur ou de la fatigue, la fonction sociale de ces propriétés mentales est au minimum d'orienter le sujet, en tant que sujet social conscient, à agir pour réduire l'intensité du désagrément (douleur, fatigue). Un autre exemple peut illustrer cela. Si un enfant dit « j'ai froid », personne ne lui dira « tu n'as pas froid car il fait 20° ! », l'expression du ressenti de froid est un moyen, pour agir (l'enfant est-il malade ? Fatigué ?...).

L'instrument de mesures subjectives est adapté à l'objet mesuré : une propriété mentale, qui n'existe que sous la forme sociale de ses manifestations. Ces manifestations passent le plus souvent, dans un protocole de mesure, par une verbalisation auto-administrée (ou hétéro-administrée), avec le biais potentiel d'une « distorsion » du discours. Mais en tant que discours ayant une fonction sociale, l'expression verbalisée révèle un état mental précis. C'est le ressenti, c'est-à-dire l'existence en tant que fonction sociale, que nous mesurons. Si on peut critiquer l'usage abusif de mesures quantitatives (et là je rejoins Vautier), l'idée que je veux défendre ici est que de telles mesures quantitatives sont potentielles. Je considère que l'on peut appréhender la douleur, la fatigue ou le ressenti de froid par exemple comme pouvant se représenter sur une échelle quantitative pour avoir une objectivation de l'intensité de l'état du sujet dans l'objectif d'agir en conséquence. Vautier considère que « *Si, par exemple, la personne énonce "7 cm" hier et "6 cm" aujourd'hui, il est possible qu'elle dise "j'ai un peu moins mal aujourd'hui qu'hier" dans le langage du clinicien qui la fait s'exprimer par déplacements du curseur de l'échelle. Mais il est impossible de déduire de ces énoncés qu'elle a objectivement moins mal* ». La divergence, entre Vautier et moi, est sur le terme « objectivement ». Pour moi « l'objectivité » est la réalité vécue, c'est-à-dire cette propriété mentale en tant que objectivée/conscientisée. De ce point de vue, Durieux, Bruxelles,

Savignoni, & Coste (2001, p. 575) considèrent que « *L'autoévaluation de la douleur, qu'elle soit réalisée à l'aide d'une échelle numérique, d'une échelle visuelle analogique ou d'une échelle verbale simple est une mesure pertinente et facile à réaliser de la douleur ressentie par le patient* ».

Dans l'exemple de Vautier, nous pouvons dire, avec une potentielle erreur de mesure dont je parlerai par la suite, que la personne a sûrement « un peu moins mal » si elle dit passer de 7 à 6 sur une échelle de douleur. Mais si elle exprime hier une douleur de 7 et aujourd'hui de 1, nous concluons, et Vautier aussi certainement, que la personne a « objectivement » moins mal. Et c'est bien dans une démarche pratique qu'une telle mesure peut s'opérationnaliser, apportant une information « objective » sur le ressenti des patients permettant d'actionner des pratiques médicales (voir par exemple Durieux et al., 2001) pour l'usage pratique des échelles de douleur).

2.3.6.4 Mesures *subjectives* et mesures *perceptuelles*

Leplège & Coste (2002) différencient la *mesure perceptuelle*, se positionnant « d'un point de vue », et la *mesure subjective*. Leplège (2001) différencie ces deux approches de la sorte : « *Cette distinction repose sur l'affirmation qu'il n'est pas suffisant de dire que les patients répondent à des questions élaborées par des experts pour affirmer que leurs réponses reflètent le point de vue des patients. Avec les mesures de qualité de vie, la médecine a donc entrepris de changer de perspective pour adopter celle du patient, ce qui l'a conduite à substituer à son objet traditionnel (la santé) la qualité de vie de ses patients* » (p. 33). Je ne suis pas certain qu'il faille différencier mesure *subjective* et mesure *perceptuelle* car c'est bien une objectivité subjective que nous mesurons, réelle mais à comprendre dans sa fonction interactive, avec une mesure *subjective*.

Donc, à mon sens, toute mesure *subjective* est une mesure *perceptuelle*, car toujours posée d'un point de vue. Comme le dit Leplège (2001) : « *Il n'est donc pas de mesure de la santé et de la qualité de vie qui ne s'opère à partir d'un point de vue* » (p. 52), mais ceci est vrai pour toutes les mesures subjectives. Cependant, je m'accorde avec ces auteurs sur l'orientation de cette affirmation, qui rompt avec une prétention d'une vision médicale objective à partir de mesures subjectives (mesures orientées vers des dimensions techniques vs des mesures où « *ce sont les sujets qui commandent l'évaluation* », Leplège, 2001, p. 55).

Mais je crois que fondamentalement, ce n'est pas tant l'instrument qui différencie cette position, que la position du praticien qui se leurre sur son instrument de mesure. La mesure ne

mesure pas une réalité médicale, car il y toujours une intersubjectivité dans une mesure, liée au contexte et enjeux de cette mesure. Je me retrouve dans ce cadre défini par Leplège (2001, p. 25), pragmatique car relationnel : « *L'objectif de ces recherches est pragmatique. Les théoriciens s'emploient à préciser les termes d'un débat rationnel, appuyé sur des observations empiriques et des mesures, qui permettent d'améliorer la qualité de vie des patients par un surcroît de pertinence en matière d'interventions médicales.* ».

2.3.6.5 Etat pur et impur ?

En accord avec Vautier, l'expression d'une douleur (ou d'une fatigue) n'est qu'un acte de langage, mais comme tout acte de langage il s'inscrit dans un contexte d'interaction social et d'expression de soi dans ce contexte. Je me retrouve donc en désaccord avec Vautier, mais aussi partiellement avec Putnam. Putnam (1984) écrit « *un état mental est pur si sa présence ou son absence ne dépend que de ce qui se passe à l'intérieur du sujet* » donc qui ne dépend pas « *de quelque chose d'extérieur à mon corps et mon esprit* » et de prendre « *la douleur* » comme exemple « *la douleur est donc un état mental pur mais le savoir est un état mental impur* ».

Putnam différencie l'état mental et la connaissance d'un état mental. J'ai du mal comprendre d'un point de vue pragmatique, c'est-à-dire en considérant les objets n'existant que par l'interaction (comme Putnam le considère), ce que Putnam peut considérer être « un état pur » car il n'y a jamais d'état mental épuré d'un contexte ; l'être est sujet social, donc il n'existe jamais « en lui », son être ne peut pas dépendre uniquement « *que de ce qui se passe à l'intérieur du sujet* » c'est-à-dire indépendamment de son être-sujet-social. De plus, je ne vois pas ce que peut être une « douleur pure » différente de « la connaissance de la douleur ». La douleur, en tant que propriété mentale, est nécessairement « connaissance » de douleur car la douleur est « manifestation de douleur » car « *ce qui nous intéresse, c'est bel et bien ce que ressent le sujet* » (Falissard, 2008b). De même pour la fatigue qui ne peut être, en tant que propriété mentale ayant une fonction sociale, que connaissance de cette fatigue. En ce sens, je rejoins Searle contre Putnam : « *Certains états mentaux, tels que la douleur, sont ontologiquement subjectifs* » (Searle, 1996).

2.4 La pratique de la mesure

Je n'ai développé pour le moment que des arguments en faveur d'une faisabilité de la mesure de propriétés mentales. L'opérationnalisation de mesures va renvoyer à d'autres problèmes. Je ne développe ici que les problèmes liés aux mesures quantitatives, car elles sont au cœur de la discussion critique dans la littérature académique.

2.4.1 Le problème du modèle statistique en psychologie

Une représentation sur une échelle numérique ne prétend pas refléter une vraie propriété au sens d'une représentation d'une réalité, mais bien comme une représentation objectivée. Mais son opérationnalisation va user de modèles statistiques qui ne sont pas sans poser des questions.

2.4.1.1 Lois et langage logico-mathématique

Comme le dit Falissard (2008b) « *La mesure est donc avant tout une commodité, un outil qui permet de mieux se représenter les faits que l'on étudie [...] faciliter la compréhension de phénomènes trop complexes pour être appréhendés directement.* ». L'homme catégorise et a besoin de construire un cadre formel de cette catégorisation. Le cadre formel des mathématiques, et le formalisme logique associé, est une manière usuelle en science pour représenter la réalité, ou une relation théorique entre différents objets. Sherry (2011) considère « *our ontology requires mathematical objects because they are indispensable for science.* ». Putnam (1998) a écrit un petit ouvrage dédié à la logique mathématique. Putnam pose que l'usage d'une représentation mathématique est indispensable en physique car on ne peut construire de lois sans recours au langage logico-mathématique, c'est-à-dire sans recours aux nombres, aux fonctions, aux ensembles, ... De même Vautier (2012) cite Granger : « *La connaissance scientifique de ce qui relève de l'expérience consiste toujours à construire des schémas ou modèles abstraits de cette expérience, et à exploiter, au moyen de la logique et des mathématiques, les relations entre les éléments abstraits de ces modèles, de façon à déduire finalement des propriétés correspondant avec suffisamment de précision à des propriétés empiriques directement observables* ».

Mais ce qui est « vrai » pour la physique, à la recherche de lois, l'est-elle pour la psychologie ? Le langage mathématique en psychologie renvoie à des questions relatives à la

logique mathématique et son efficence pour la psychologie. Granger (2002) écrit : « *Les sciences de l'homme se trouvent de par leur propre nature placées devant un dilemme. Ou réduire totalement les significations à des sens strictement mis en forme, au risque de voir s'évanouir la spécificité de leur objet, et dégénérer en exercices logico-mathématiques. Ou introduire massivement et naïvement les significations comme telles, et, perdant toute possibilité de construire vraiment des modèles abstraits, se muer, dans le meilleur des cas en herméneutique philosophique, en vaticinations mythiques dans le pire.* » (p. 272).

Il y a en effet un grand risque en usant de l'écriture logico-mathématique de glisser vers la recherche de « lois » en psychologie au sens lois déterministes. Or c'est l'absence de telles lois déterministes qui autorise certains auteurs à considérer la psychologie comme non-scientifique. J. Kim (2009) considère que « *la recherche de lois pour constitutive de la nature même de la science [...] faute de lois psychologiques, la psychologie ne saurait être une science [...] [les chercheurs en psychologie] ne produisent pas des théories psychologiques, des systèmes englobants et intégrés de lois générales exactes, formulés dans un vocabulaire proprement théorique, à partir desquels les phénomènes mentaux pourraient être expliqués et prédits* ». J. Kim se base sur Davison qui positionne « *la psychologie comme une philosophie* » et non comme une science, car pour Davidson il n'y a pas « *de lois déterministes strictes* » (cité par J. Kim pp.21-22). Cette question du caractère scientifique de la psychologie renvoie à l'objet de la science. Je suis en accord avec Putnam (1998) : « *il est difficile de croire qu'il puisse exister une chose telle que " le but de la science" »* (p. 67). Même si l'objet de la recherche en psychologie n'est pas des lois déterministes, je considère que la psychologie est pour autant dans une démarche scientifique en posant un cadre de discussion collectif pour faire progresser la connaissance, au sens où Bourdieu (2001) pose ce collectif comme communauté scientifique.

Juhel (2013a, p. 13) définit l'objectif de la psychologie comme « *de rendre compte des mécanismes psychologiques qui sous-tendent les diverses formes d'invariance du comportement humain.* ». Je peux adhérer à cette définition si on comprend qu'en psychologie, comme dans tout champ scientifique, il y a « *de l'ordre, une logique, bref quelque chose à comprendre dans le monde, y compris dans le monde social* » (Bourdieu, 2001). Mais à contrario l'usage de terme « *d'invariance* » peut renvoyer me semble-t-il à des formes de « lois déterministes » (cette formulation peut entraîner l'idée que l'objectif serait de rendre compte des formes déterministes, sans variance, du comportement humain). On retrouve un glissement vers un déterminisme aussi chez Vautier (2015g) qui définit la psychologie comme « *la découverte de principes psychologiques généraux c'est-à-dire*

valables pour tout système psychologique et falsifiables (i.e., tels que l'observation d'un cas particulier peut en droit falsifier la théorie) », qui me semble renvoyer aussi vers une recherche de lois déterministes. A mes yeux, si la psychologie se contente comme objectif de rendre compte « d'invariance » ou de « principes valables pour tout système psychologique », le champ est faible car il ne peut s'agir que de lois déterministes rares en psychologie. Si on entend « formes d'invariance » dans un sens plus souple (« forme »), je peux adhérer à la définition de Juhel.

La psychologie est, en effet, en tension entre une forme de déterminisme et une forme d'indéterminisme. Il y a du sens dans les propriétés mentales, donc nous cherchons des règles de déterminations... mais sans considérer que ce sens soit régi comme pour les lois des sciences de la nature, ne serait-ce que par la nécessité de renégocier les objets dont la psychologie discute (des propriétés mentales inscrites dans une pratique sociale). Je considère, sans vouloir définir précisément l'objet de la psychologie, que nous recherchons plutôt des « tendances », des « propensions », de certains mécanismes psychologiques ; sans affirmer que de telles relations soient « invariantes », ou qu'elles soient des « principes généraux », mais bien des relations « en moyenne » opérantes si on les considère inscrites dans une pratique à laquelle le chercheur appartient. C'est-à-dire des lois « probables » (au sens statistique) et non « déterministes » de relations entre propriétés mentales considérées comme appréhendées relativement à un contexte et enjeu (donc bien en rupture avec une prétention de déterminer des relations entre propriétés qui seraient valides indépendamment de notre pratique sociale). En accord avec Rouxel (2015) « *Les psychologues, quelle que soit leur orientation, ne peuvent qu'être d'accord avec l'idée selon laquelle les personnes possèdent des dispositions à se comporter de façon consistante [...] à minima dans certaines classes de situations* » (p. 127). On parle de « consistance » ici et non « d'invariance » ou de « principes valables pour tout système psychologique ».

Le recours au langage logico-mathématique me paraît une nécessité en psychologie dans certaines situations, si ce recours est inscrit dans cette démarche claire de refus de chercher des « lois », mais bien d'user de modélisation dans le sens de faire progresser la connaissance pour comprendre des mécanismes « probables », c'est-à-dire user du langage logico-mathématique pour poser des modèles statistiques et non des modèles déterministes.

2.4.1.2 Limites du langage logico-mathématique

Comme le dit Putnam (1998), dans un formalisme mathématique « *Ce qui est vrai ce n'est pas l'énoncé [...] mais ce que signifie l'énoncé* » (p. 21). Il doit donc y avoir un retour vers le « sens » d'une formalisation logico-mathématique en psychologie car tout le cadre de référence (catégories, mesures, ...) est inscrit dans une pratique précise (contexte, enjeu). Je peux reprendre les mots de Merleau-Ponty (1960/2001, p. 127) en l'appliquant à la psychologie : « *Si l'objectivisme ou le scientisme réussissait jamais à priver la sociologie de tout recours aux significations, il ne la préserverait de la "philosophie" qu'en lui fermant l'intelligence de son objet. Nous ferions peut-être alors des mathématiques dans le social, nous n'aurions pas la mathématique de la société considérée. Le sociologue fait de la philosophie dans toute la mesure où il est chargé non seulement de noter les faits, mais de les comprendre.* ».

Donc clairement, le recours à un langage logico-mathématique ne doit pas annihiler le sens que les mathématiques ne donnent pas. Armatte (2005) considère, avec raison je pense, qu'un modèle mathématique est un formalisme logico-mathématique qui est une abstraction extraite d'une réalité ; et donc le terme « modèle » renvoie à une vision normative qui doit être appréhendée avec du recul en psychologie car tout cadre normatif oublie les singularités. Armatte cite Suppes : « *une théorie est une entité linguistique faite d'énoncés alors que les modèles sont des entités non linguistiques dans lesquelles la théorie est satisfaite.* », et donc avec une confusion potentielle, voire inversion de sens entre « modèle » et « théorie ». Donc l'objet d'un modèle en psychologie doit clairement être compris comme un support pour extraire du sens, passer du « modèle » à la « théorie » par le « sens », mais non dans une vision normative et déterministe.

C'est dans ce rapport psychologie/langage-logico-mathématique/sens que l'article de Falissard, Révah, Yang, & Fagot-Largeault (2013) me semble très intéressant. Ces auteurs cherchent à repositionner la taxinomie usuelle qui oppose les études qualitatives aux études quantitatives. Une nouvelle taxinomie est proposée basée sur deux axes différents : 1/ le langage symbolique de représentation des objets analysés ; et 2/ la relation entre les objets analysés. Le langage de représentation peut être « littéraire », « mathématique » ou « un mélange des deux ». La relation entre les objet peut être « herméneutique », « déterministe » ou « statistique » (au sens de relations « probables »).

TABLEAU 3 : TAXONOMIE DES APPROCHES EN RECHERCHE PSYCHIATRIQUE

		Relationships between observations are obtained from		
		A) Hermeneutics	B) Statistics	C) Equations
Symbolisation of observations	1) <i>Literary</i>	Psychoanalysis, most transcultural psychiatric studies	Computerized textual analysis	Many neurobiological mechanisms
	2) <i>Locally mathematical</i>	ABAB designs for evaluation of CBT	Epidemiology, Cognitive neuroscience, Brain imaging	Pharmacokinetics of psychotropic drugs, Lacan's "graphe du désir"
	3) <i>Globally mathematical</i>	Spin glass model of brain functioning	None at the moment in psychiatric research [thermodynamics]	None at the moment in psychiatric research [most of physics]

Objects can be represented globally with mathematics, locally with mathematics or without mathematics. Relationships between these objects can be obtained from a hermeneutic process, from equations or from statistical analyses.

Tiré de par Falissard et al. (2013, p. 5)

La psychométrie est clairement positionnée le plus souvent dans une représentation mélangeant le « littéraire » et le « mathématique » ; et l'analyse est le plus souvent « statistique ». La différence entre un modèle « mathématique » et un modèle « statistique » est caractérisé ainsi : « *Some could argue at this point that statistics are a branch of mathematics and that obviously equations are very common in this discipline (linear regression models, often used in psychiatric research, are indeed equations). The distinction between the two categories "statistics" and "equations" (B. and C.) could thus be highly questionable. The difference is that statistical procedures do not inherently provide formal, universal and deterministic relationships. In short, deterministic equations provide relationships that are supposed to be true in all instances, for everyone, while most statistical relationships are only true for an average subject, who does not exist in practice.* » (p. 7).

En reprenant ce cadre, nous pouvons considérer que la psychométrie, usant d'écriture mathématique, comme non inscrite dans une démarche fondamentalement "mathématique" mais bien "statistique" au sens donné par Falissard et al. (2013) ; c'est-à-dire cherchant non des modèles « exacts » mais bien des modèles « probables » au sens statistique du terme (vrais en moyenne en quelque sorte). Je reviendrai par la suite sur ce que ces auteurs ajoutent à la fin de l'extrait (« *true for an average subject, who does not exist in practice* ») car cela est fondamental. On peut rapprocher ce cadre défini par ces auteurs au « *style statistique* » dans l'histoire de l'argumentation scientifique tel que Hacking (1992) le discute comme trajectoire particulière de procédure de vérifications scientifique.

2.4.1.3 Tensions d'un modèle statistique

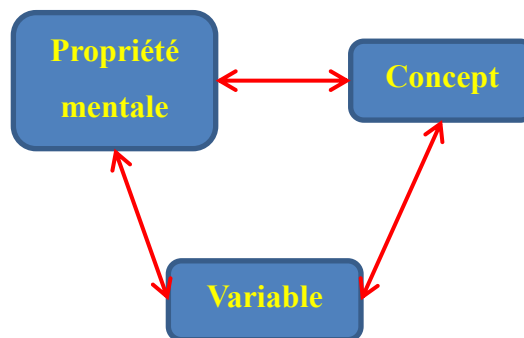
L'usage de modélisations statistique s'appuie, partiellement, sur une formalisation en langage mathématique. Cette écriture laisse souvent considérer une « objectivité » au travail quantitatif que n'aurait pas un travail s'appuyant sur des études qualitatives. Or, comme

l'explique Falissard et al. (2013), la construction du discours et de la pratique est en réalité assez proche entre les études quantitatives et les études qualitatives. Le discours d'une étude quantitative en psychologie se différencie d'un discours mathématique. Ce dernier se suffit du formalisme logico-mathématique ; les études quantitatives en psychologie introduisent littéralement le modèle et analysent les résultats empiriques de même par un formalisme littéraire et non dans un formalisme logico-mathématique. En accord avec ces auteurs, les études quantitatives ne diffèrent pas des procédures des études qualitatives car fondamentalement : « *differing degrees of mathematical representation each involve subjective and heuristic choices that impact the results and their interpretation.* » (Falissard et al., 2013, p. 8).

Il y a donc une analyse réflexive à avoir sur une approche statistique car des choix subjectifs impactent les résultats. Je m'appuie sur deux schémas pour clarifier cette position.

D'une part il faut clairement différencier une propriété mentale, le concept associé, et la formalisation mathématique de la propriété mentale.

FIGURE 13 : SCHEMA DIFFERENCIANT PROPRIETE-CONCEPT-VARIABLE

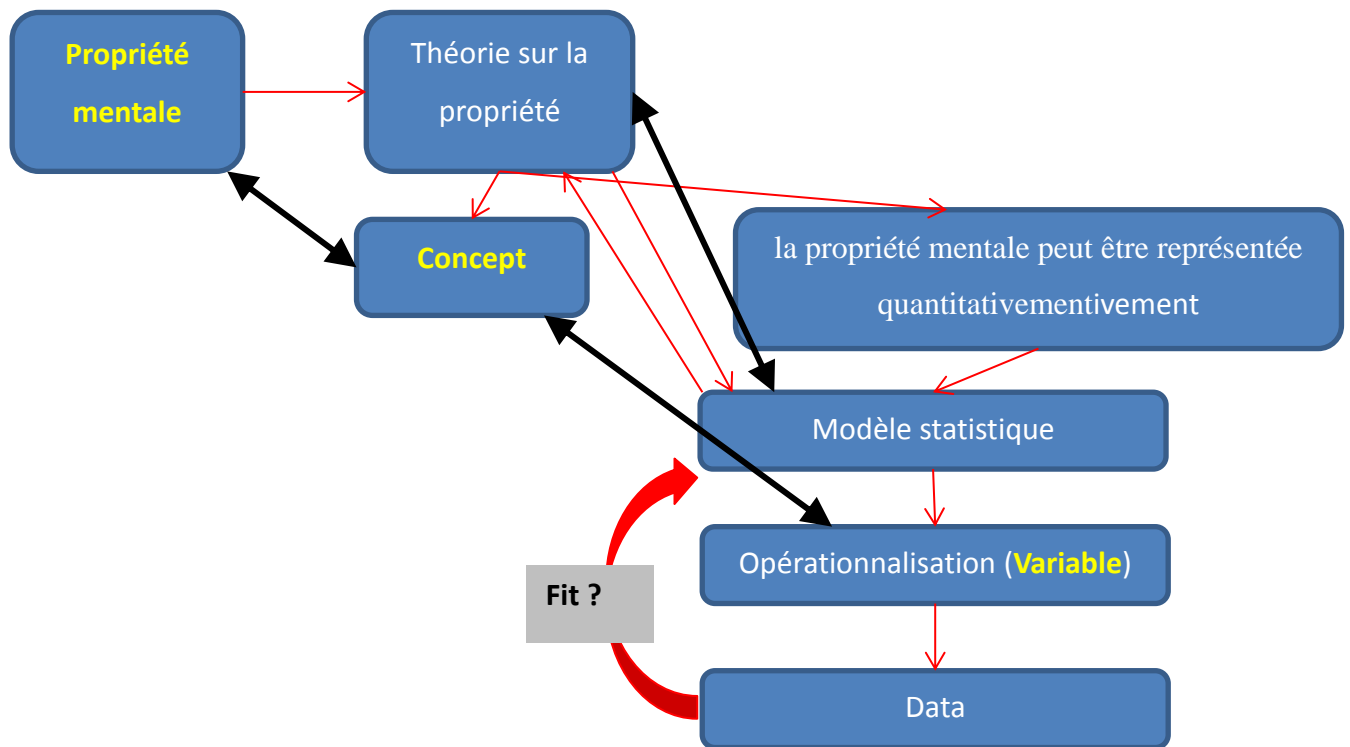


- La variable renvoie à la représentation de la propriété mentale dans un formalisme logico-mathématique ;
- le concept est la catégorisation pratique par une formalisation linguistique de la propriété mentale.

La variable et le concept sont deux formalisations différentes d'une seule et même propriété mentale ; il n'y a pas équivalence entre le concept et la variable. C'est là, de prime abord, une dérive fréquente dans la littérature que d'assimiler la variable au concept, et la variable à l'objet (Maraun & Gabriel, 2013; Maraun & Peters, 2005). Il y a une démarche de « choix » du concept et de la variable (Quels items ? Quel codage ?).

Cette tension entre propriété mentale, concept et variable va générer des tensions entre théorie et modèle statistique. Je représente dans le schéma suivant en flèches rouges les « implications » et en flèches noires les « tensions » dans une modélisation.

FIGURE 14 : SCHEMA DU PROCESSUS D'UNE APPROCHE QUANTITATIVE



- La propriété mentale étudiée renvoie à une « théorie » sur la propriété mentale inscrite dans une pratique et non comme une « vérité » en soi ;
- Cette « théorie » sur la propriété mentale (unidimensionnalité de la propriété mentale, indépendance locale, ...) permet de « conceptualiser » la propriété mentale dans un cadre pratique ;
- Cette théorie sur la propriété mentale permet de poser si la propriété mentale peut être représentée ou non sur une échelle numérique ;
- Une telle formalisation dans le langage numérique va autoriser une formalisation d'un « modèle statistique ». Ce modèle représente dans le langage logico-mathématique la « théorie », mais il n'y a pas équivalence entre les deux. Concrètement, une même théorie sur une propriété mentale peut faire l'objet de différentes modélisations. Si je prends par exemple les modèles d'équations structurelles (SEM), nous pouvons

utiliser les SEM soit par l'approche CB-SEM (*covariance based*) ou soit par l'approche PLS-SEM (*partial least square based*) qui répondent à deux modèles mathématiques différents pour un même cadre théorique (et donneront des résultats différents, parfois très proches, parfois éloignés).

- Ce modèle statistique renvoie à l'opérationnalisation²⁶ des variables, la mesure concrète pour résumer. Cette opérationnalisation d'une mesure n'est pas unique et peut être entreprise de différentes manières (choix des manifestations ; choix des mesures, sur une échelle de Likert ou autres par exemple ; choix des items ; ...). Il y a donc un choix dans l'opérationnalisation et donc pas d'identité entre la variable opérationnalisée et le concept. La variable opérationnalisée est une démarche de « représentation » pratique basée sur un langage logico-mathématique pour mesurer la propriété mentale ;
- Le modèle statistique est validé, ou non, par des critères « statistiques » (que j'ai résumé par « fit »)... avec comme nous le savons tous les problèmes liés (puissance des *Goodness of Fit*, GOF ; potentiel modèle concurrents ; ...). Donc la validation « statistique » d'un modèle n'est en rien une validation du modèle théorique.

Il y a donc tout au cours de la formalisation et de l'analyse des « choix » et des « heuristiques » qui demandent un retour réflexif. Je renvoie pour exemple au livre de Desrosières (2008a) qui discute entre autres des implications/choix liées au choix d'user de la régression logistique ou de l'analyse des correspondances par exemple (chapitre 1).

2.4.1.4 Validité d'un modèle statistique ?

Si je développe ce schéma, c'est pour pointer à mon sens les erreurs qui me semblent exister dans la littérature académique. Combien d'articles après avoir « modélisé » et « validé » un modèle statistique (par un GOF ou une p-value) prétendent sans retenu avoir « montré » (voire « démontré ») la validité du modèle théorique. Du modèle statistique à la théorie il y a l'ajout du « sens », c'est à dire que nous remontons d'un formalisme logico-mathématique (avec tout l'avantage de ce formalisme, mais aussi son incapacité à donner du « sens ») à notre conception pratique de la propriété mentale. C'est ce sens que donne le cadre théorique,

²⁶ « une théorie est operationaliste quand les termes théoriques, auxquels elle recourt, sont définis par l'énonciation de procédures expérimentales permettant de leur assigner une grandeur de les mesurer. (Drouet, 2012, p. 12)

ce que Falissard et al. (2013) considèrent comme le discours dans un formalisme littéraire et non plus logico-mathématique. Le modèle statistique, son opérationnalisation et sa validation, ne sont que des supports à l'intelligence scientifique et ne portent en eux-mêmes aucun « sens » pratique. C'est en cela que la pratique concrète dans l'écriture d'un travail de recherche use du langage littéraire pour donner ce « sens » aux résultats... comme une étude « qualitative » (cf. Falissard et al., 2013).

Je me permets de développer rapidement cette « fragilité » relative de résultats statistiques, non pour considérer qu'elles ne fournissent pas d'information, mais pour rappeler que cette information statistique ne prouve pas la validité du modèle théorique (entre autres, aucun modèle statistique ne prouve qu'une propriété mentale soit mesurable). Si je considère les deux formes qui paraissent les plus usuelles avec les variables latentes, le modèle de Rasch et les SEM, nous pouvons souvent lire cette hérésie (je reprends le terme de Maraun) de considérer qu'un modèle statistique, validé par un indicateur statistique (GOF), valide le modèle théorique. Pour le modèle de Rasch, P. Barrett (2005), Borsboom & Scholten (2008), Michell (2004), Scholten (2011b), Sijtsma (2012b) par exemple illustrent sur des exemples précis que des données issues d'une propriété mentale non quantitative peuvent valider empiriquement un modèle de Rasch. Pour les SEM, de même, bien des travaux montrent la faible puissance des indicateurs statistiques (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Fan, Thompson, & Wang, 1999; Steiger, 2007) et le fait que des modèles empiriques valides statistiquement sont néanmoins erronés relativement aux cadres théoriques de données générées (Jarvis et al., 2003; MacCallum & Browne, 1993). Sijtsma (2012b, p. 800) écrit pour le modèle de Rasch, mais cela peut s'étendre aux SEM et autres modèles avec variables latentes : « *however, an IRT model cannot “know” that the operationalization was incorrect and thus cannot reveal this [...] However, the Rasch model is a statistical model, and I contend that whether it refers to the real world depends on whether its formal structure matches the theoretical structure of an attribute.* ».

Je peux illustrer par un autre exemple la tension entre modèle théorique et validation empirique. L'alpha de Cronbach est encore aujourd'hui souvent utilisé pour valider une échelle de mesure d'une propriété mentale en tant que validation de l'unidimensionnalité et fiabilité de cette échelle (*reliability*). Or l'usage de l'alpha de Cronbach semble en contradiction avec ses qualités mathématiques. Je ne développe pas ce point mais je renvoie à quelques articles qui pointent l'efficacité relative de cet indicateur et sa nature « mathématique » vs son « usage » : Green & Yang (2009); Rouquette & Falissard (2011); Sijtsma (2009a, 2009b).

2.4.2 Le problème des probabilités

Les modèles avec variables latentes sont des modèles « statistiques ». Ils cherchent à caractériser des relations probables et non-déterministes ; de plus ils se caractérisent par l'introduction des erreurs de mesure et d'erreurs systématiques (Sijtsma, 2012a, 2012b). Cela oblige les modèles avec variables latentes à poser des modèles ayant recours aux variables aléatoires, donc renvoyant aux probabilités. L'usage de probabilités couvre une discussion large que je ne prétends pas développer, je discute juste du problème posé par certains auteurs sur la légitimité de considérer des probabilités comme des réalités.

2.4.2.1 Les deux formes des probabilités

Une probabilité peut être comprise dans deux sens différents avec une propriété mentale, en référence à ce que Hacking (2002) développe : *« Il est notoire que cette probabilité qui émerge si soudainement a les deux visages de Janus. D'un côté, elle est statistique et s'applique aux lois stochastiques des probabilités aléatoires. De l'autre, elle est épistémique et concerne l'évaluation des degrés auxquels il est raisonnable de croire en la vérité de propositions n'ayant rien de statistique. »* (p. 38). Je reprends les mots de Drouet (2012) pour différencier « probabilité ontologique » et « interprétation épistémique » : *« il y a des probabilités parce que certains événements sont, effectivement, aléatoires. Les probabilités et leur valeur sont alors des caractéristiques du monde physique. Nous parlerons, en ce sens, d'interprétations ontologiques des probabilités. [...] Aux interprétations ontologiques des probabilités, on oppose généralement les interprétations épistémiques. Selon ces interprétations, les probabilités ne sont pas des caractéristiques du monde physique, mais ont à voir avec notre connaissance. Plus exactement, selon les interprétations épistémiques, la probabilité d'une proposition mesure sa crédibilité dans certaines circonstances. »* (p.3)

En psychologie, il y a fondamentalement un problème épistémique, qui est notre manque de connaissance des processus mentaux. Mais il y a aussi un problème aléatoire potentiel qui renverrait à la non-prédictibilité du comportement humain, au « libre arbitre » en rupture avec des lois déterministes comme le supposent un certain nombre d'auteurs. Il est à impossible de différencier les deux dans le cadre de nos connaissances actuelles. De plus, même le caractère épistémique de l'aléa d'une propriété mentale pose la question de notre incapacité future à dépasser cet aléa épistémique, comme je l'ai discuté dans la première partie de ce travail, car l'émergence d'une propriété n'est pas réductible aux processus qui l'ont générée (même si cela se fait par des lois logiques, donc même indépendamment du « libre arbitre » en potentiel

rupture avec des lois logiques). Belis (1995) considère : « *Modéliser l'incertitude — voire l'ignorance — est une tâche ardue, vu qu'il est très difficile de séparer l'incertitude subjective (due à l'insuffisance de nos connaissances) de l'indétermination objective qui peut exister dans la réalité environnante* » (p. 200).

Mais l'usage des probabilités en psychologie est surtout justifié par les erreurs de tout protocole de mesure en psychologie, comme déjà expliqué. Donc, au-delà de sa non-prédictibilité des propriétés mentales, il y a de fait des erreurs associées qui obligent le recours à une formalisation sous forme de variables aléatoires. Une question centrale qui est donc posée est la « réalité » de ces probabilités associées à des propriétés mentales. Si de telles probabilités ne pouvaient prétendre à être des « réalités », les modèles avec variables latentes seraient de simples fictions pour bien des auteurs. Pour la suite de mon développement, cela ne me paraît pas utile de devoir différencier l'approche fréquentiste de l'approche bayésienne des probabilités car, comme le pose Scheemaekere (2009), cela n'est pas impératif pour la discussion posée ici.

2.4.2.2 Probabilités et propensions

Les probabilités sont souvent questionnées relativement à des « propensions ». Pour Popper « *les propensions sont des « propriétés non observables du monde physique [...] douées de tendances ou de dispositions à se réaliser* » (cité par Belis, 1995). Concrètement, en reprenant Drouet (2012) « *Selon l'interprétation propensionniste, les probabilités mesurent ce que Popper appelle des "propensions". Il s'agit de tendances de certains dispositifs physiques à réaliser certains événements possibles. Ainsi, pour le propensionniste, dire que la probabilité d'obtenir pile à l'issue du lancer d'une pièce équilibrée vaut 1/2, c'est dire que le dispositif de lancer de la pièce possède une propension à réaliser l'événement pile et que cette propension est mesurée par le nombre réel 1/2* » (p 1).

Cette notion de propension renvoie à une nature ontologique des probabilités en tant que caractérisant des réalités de phénomènes. En m'appuyant sur Drouet (2012) : « *le propensionnisme poppérien prend pour point de départ l'idée selon laquelle "la situation physique globale" à un instant donné inclut des tendances à réaliser les événements qui sont possibles dans la situation considérée. Ce sont ces tendances que Popper appelle "propensions". Les propensions sont des entités comparables aux forces newtoniennes, à deux titres : elles sont à la fois impossibles à observer directement et douées d'une réalité physique : "Les propensions, comme les forces d'attraction newtoniennes, sont invisibles, mais, comme*

ces dernières, elles peuvent agir : elles sont effectives, elles sont réelles". Plus précisément, ce passage fait apparaître que la réalité d'une propension consiste dans son activité en vue de la réalisation d'un événement possible ou, plus généralement, dans sa capacité à engendrer un (ou des) phénomène(s) observable(s). La réalité des propensions est susceptible de degrés et elle est mesurée, pour chaque propension, par la probabilité de l'événement qu'elle tend faire advenir : "dans l'interprétation propensionniste, la probabilité d'un résultat est (...) la mesure de l'inclination de l'état de choses actuel à produire ce résultat". Selon cette conception, les probabilités, à l'instar des propensions qu'elles mesurent, sont des caractéristiques de la situation physique considérée. Par ailleurs, l'inscription des propensions dans la réalité physique implique que les événements qu'elles tendent à réaliser sont inscrits dans l'espace et dans le temps : ce sont des événements singuliers plutôt que des événements génériques. » (p. 6). Pour Popper « La propension n'est pas inhérente aux choses prises individuellement [...] elle est une propriété relationnelle d'un arrangement expérimental [...] La propension est donc un concept relationnel » (Scheemaekere, 2009, p. 86). Pour Popper il semble y avoir une assimilation entre « propension » et « probabilité » : « Les probabilités doivent être physiquement réelles [...] elles doivent être des propensions physiques, des propriétés relationnelles abstraites du contexte physique, comme les forces newtoniennes » (cité par Scheemaekere, 2009)

Or cette assimilation d'une « propension » à une « probabilité » a été critiquée. Kyngdon (2008a) considère qu'il faut différencier une « propension » d'un objet et une probabilité. Une propension d'un objet est un effet de l'objet plus que sa structure intrinsèque ; mais au-delà, les propensions ne respectent pas les lois de probabilités et on ne peut donc pas les considérer comme des équivalents. Kyngdon renvoie à Humphrey et les paradoxes qu'il a soulevés. Humphreys (1985) discute de la différence entre « propension » et « probabilité ». La réalité ne respecte pas le théorème de Bayes (probabilité conditionnelle) : par exemple poser la probabilité d'avoir un cancer est dépendant de la condition de fumer; par contre la probabilité de fumer est indépendante d'avoir ou non un cancer (non détecté). Humphrey prend un autre exemple : la probabilité conditionnelle qu'une vitre se casse si on jette une pierre et la probabilité de jeter une pierre si une vitre se casse. Ces exemples discutent des probabilités conditionnelles formalisées mathématiquement (objets abstraits) et des « propensions » (objets réels) qui ne respectent pas les lois de la probabilité car il n'y a pas de symétrie dans les propensions, à la différence des probabilités. Une vitre a une propension à se casser quand une pierre est lancée dessus, une pierre n'a pas de propension d'avoir été

lancée quand une vitre se casse. La théorie des probabilités est une théorie inscrite dans la logique mathématique et donc ne peut pas être considérée comme réelle.

2.4.2.3 La probabilité comme mesure

Si je considère que la critique de Humphreys renvoie à de vraies questions sur la nature réelle des probabilités, je suis par ailleurs en opposition avec ses arguments contre le formalisme des probabilités qui seraient à critiquer, car pour Humphreys la théorie des probabilités ne représente pas bien les propensions. Un formalisme représentatif d'une réalité ne peut prétendre, par toutes ses combinaisons, être un représentant de la réalité. Pour exemple, si les mots sont des représentants d'une réalité : les mots « corbeau » et « blanc » définissent des réalités (disons des catégories objectivées) ; cela n'oblige pas à ce que « corbeau blanc » soit une représentation d'une réalité (ou d'une catégorisation de la réalité objectivable). De même, les mathématiques usent de nombres complexes dans la formalisation de la réalité sans prétendre que l'imaginaire « i » existe (Quine : « *no measurement could be too accurate to be accommodated by a rational number, but we admit the [irrationals] to simplify our computations and generalizations* » cité par Sherry, 2011).

Ainsi si on considère les probabilités comme un formalisme logique de représentation de certaines propriétés (une mesure d'une propension), cela ne veut pas dire que toute construction humaine à partir de ce formalisme doit être une représentation d'une réalité. Ce sont des outils nécessaires pour formaliser la réalité dans un langage mathématique, et non la réalité elle-même. Si je reprends l'exemple de la vitre et de la pierre, nous avons :

A = « vitre cassée » ; B = « pierre lancée »

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \quad (4)$$

et

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \quad (5)$$

Bien évidemment que ces probabilités conditionnelles ne peuvent prétendre « représenter » la réalité car la probabilité que la pierre ait été lancée sachant que la vitre est cassée n'est pas une « propension » de la pierre. Par contre cette probabilité n'est pas contre-intuitive au niveau du formalisme logique des probabilités (on peut poser un modèle probabiliste et estimer empiriquement la fréquence du nombre de pierre ayant cassé une vitre vs le nombre de vitres cassées sans jet de pierre par exemple). De ce fait, autant on peut en effet se démarquer de Popper et considérer les probabilités comme « non-réelles » car simples

représentations ; autant le système de représentation semble opérant pour « mesurer » des propensions et donc se démarquer de Humphrey.

En fait, on retrouve ici la tension entre réalité et représentation. La propension est réelle, la probabilité est une représentation de cette réalité, elle n'est pas réelle ; tout comme le « sexe » est une variable qui est une représentation du réel, mais n'est pas équivalente à la réalité. Je me positionne donc en critique de Popper qui voyait les probabilités comme des réalités. De plus, comme le rappelle Scheemaekere (2009) l'axiomatisation de Kolmogorov assimile la probabilité à une mesure (au sens de la théorie de la mesure de Lebesgue par exemple). Dans les probabilités, on retrouve la démarche de la RTM pour la représentation d'une mesure d'une propriété, et non l'identité entre les deux.

La question centrale est donc avec des probabilités de savoir si nous pouvons construire un isomorphisme entre les propensions et leur représentation par des probabilités. Formellement, les probabilités suivent les axiomes de la mesure (entre autre additivité) ; mais la représentation par une probabilité oblige à questionner empiriquement si les propensions sont sur un continuum, donc on revient au problème déjà discuté de la propriété mentale et sa mesure tel que développé antérieurement.

2.4.2.4 L'approche pragmatique des probabilités

Cela renvoie à la conception pragmatique de la probabilité qui est subjective (Scheemaekere, 2009). On peut considérer qu'une probabilité est un moyen pour l'homme de représenter des réalités (des propensions), donc liée aux caractéristiques du moment. Scheemaekere (2009, p. 65-66) considère que les probabilités approchées par le courant pragmatiste se base sur « *Le refus de faire de la probabilité une entité objective autonome qui existe indépendamment d'un jugement personnel [...] la théorie logique revient à attribuer un contenu réel à des expressions logico-mathématiques qui sont purement formelles ou symboliques* ». Scheemaekere (2009, p. 79) cite Peirce: « *afin d'interpréter le concept de probabilité du point de vue pragmatisme, il nous faut nous demander ceci : "quand nous disons que la probabilité d'un évènement a une certaine valeur p , que voulons-nous dire ?". D'après la maxime du pragmatisme, il faut, ensuite, se demander quelle différence pratique découle du fait que cette valeur soit p , ou non. Ensuite, nous devons examiner comment les probabilités sont employées dans les affaires pratiques* ». Ainsi pour Peirce, « *la probabilité est un "would be", c'est-à-dire une propension, une propriété à long terme, comparable à une*

habitude. C'est un caractère dispositionnel ». Il y a donc une forme de convergence entre Popper et Peirce.

Mais si on distingue probabilité épistémique et probabilité physique, les deux sont liées dans la démarche pragmatique, c'est une subjectivisation par les probabilités d'une propension réelle. Scheemaekere (2009) « *C'est donc le rapport entre les structures mathématiques et la réalité qui pose question* » (p. 107), or la réalité est une objectivation « pour nous », formalisée dans un langage symbolique. Donc la probabilité mesure une propension objectivée, mais renégociable car dépendant du contexte, comme le dit Scheemaekere (2009).

2.4.2.5 Conclusion

En conclusion, tout comme la mesure en générale, associer des probabilités à une propriété mentale est une démarche qui doit se considérer comme pragmatique, sans opposer « réalité » et « représentation » car la réalité n'est objectivée que sous forme de représentation. Les probabilités vérifient, par le formalisme associé, les axiomes de la mesure, donc rien ne semble interdire l'usage de modèles avec variables aléatoires pour modéliser notre objectivation de la réalité autant que la propriété mentale en question puisse rationnellement être considérée comme représentable sur un continuum numérique, et donc sur un continuum de probabilités. L'usage d'une représentation par un modèle statistique faisant appel aux probabilités (ou variables aléatoires), en tant que représentation d'une réalité, me semble parfaitement légitime.

2.4.3 Le problème de la falsifiabilité

Vautier (2015f) considère que le modèle probabiliste maquille les « anomalies » des données relativement au modèle théorique (concrètement, des items considérés comme hiérarchisés verront des personnes réussir des items « plus difficiles » et échouer à des items « plus faciles », donc contradictoire avec la théorie) : « *L'intuition quantitative est sauvegardée mais c'est au prix d'un renoncement à la connaissance théorique.* ». En conséquence, les modèles avec variables latentes ne seraient pas testables pour Vautier (2011) : « *We shall see that an empirical law consists precisely of stating an empirical impossibility, i.e. a partially deterministic falsifiable statement.* », et Vautier et al. (2014) : « *Consequently, individual observations cannot play the role of falsifying cases in this paradigm* ».

Cependant, comme Guttman le considérerait « *perfect scales are not to be expected in practice* » (cité par Engelhard, 2008, p. 165). En effet, la psychologie n'est pas une science dure dans

le sens où les théories ne renvoient que très rarement à des modèles déterministes, mais à des “propensions” comme je l’ai discuté, modélisés par des modèles statistiques. La falsifiabilité des modèles doit donc s’inscrire dans un tel cadre non-déterministe mais en relation avec le cadre nécessaire de la psychologie (modèles statistiques).

2.4.3.1 Le problème de la falsifiabilité

Une des critiques centrales (Vautier par exemple) est sur la non- falsifiabilité possible des modèles de mesure car l’introduction des erreurs de mesure permet de justifier toute modélisation. De plus, sur le plan des résultats empiriques, rien ne permet de valider un modèle vs un modèle concurrent, c’est le fameux problème des modèles « équivalents » par exemple avec les SEM (MacCallum, Browne, & Sugawara, 1996). Bartholomew, Deary, & Lawn (2009) montrent d’une façon plus générale qu’une matrice de corrélations peut être comprise comme déterminée par une dimension unique ou différents facteurs, sans pouvoir déterminer un “meilleur” modèle au sens statistique du terme (GOF ou autres) : « *This means that there is no statistical way in which we can distinguish between the two types of model on the basis of their covariances alone [...] the main point of the exercise has been to show that there is a broad family of models, including both the familiar factor model and the various versions of the bonds model, which cannot be distinguished by any statistical means.* » (p. 573-574). Or la seule validité des tests est souvent leurs corrélations avec d’autres phénomènes étudiés, qui a donc une puissance faible (Meehl, 1990). McGrath (2005a) montre que pour des dimensions proches, différents tests donnent ainsi des résultats peu corrélés et il donne différents exemples de tests douteux, usuellement utilisés, car la validité croisée des tests est faible empiriquement. Cela renvoie-t-il à une impasse scientifique, les modèles de mesure s’appuyant sur des modèles statistiques ?

2.4.3.2 Critique de la falsifiabilité « dogmatique »

De prime abord et d’une façon générale il est contestable que la falsifiabilité, dans le sens strict poppérien, soit au cœur de tous les développements scientifiques. Putnam (1984) notait « *la théorie de l’évolution par sélection naturelle n’est pas fortement falsifiable ; elle ne fait pas de prédictions bien définies qui permettraient de la réfuter. Nous n’acceptons pas la théorie de l’évolution parce qu’elle a survécu à un test poppérien, mais parce qu’elle fournit une explication plausible d’une masse de données, parce qu’elle a suggéré un ensemble de nouvelles théories, parce qu’on a pu la rattacher aux développements de la génétique et de la*

biologie [...] » (p. 218). De plus, comme le note Feyerabend (1975/1988), une théorie n'est jamais en « vraie » au sens poppérien car il y a toujours des contre-exemples (même pour les théories les plus acceptées), ou des « anomalies » (Lakatos, 1970). De plus, Feyerabend (1975/1988) discute que, le plus souvent, le remplacement d'une théorie par une autre ne se fait pas de la façon poppérienne, par falsifiabilité indubitable, car le cadre théorique de Galilée par exemple ne répondait pas à des résultats empiriques (un certain nombre de contre-exemples ont été proposés à Galilée qu'il ne pouvait réfuter sur la base de son cadre théorique, voir tout le développement de Feyerabend, 1979), mais par son opportunité potentielle pour la science ; de même Lakatos (1970) illustre le même problème à partir de la théorie de Newton qui avait à son époque beaucoup de contre-exemples.

La « falsifiabilité » stricte comme cœur de toute démarche scientifique est une vision trop rigide. Tambolo (2015) par exemple développe les évolutions des positions respectives de Popper ou Feyerabend sur ce sujet ; et la perception de Popper comme positionnée sur une théorie testable sans recul est fautive. Cela renvoie à Kuhn (1970) car, si l'histoire scientifique était ce chemin simple, fait de théories puis de falsifiabilité indubitables, nous ne comprendrions pas pourquoi à chaque étape il y a débat, exemples et contre-exemples, différentes théories en concurrence (voir tout le développement sur ce point dans Callon & Latour, 1982/2013). L'histoire des confrontations scientifiques de fait est basée sur un cadre de comparaison/ falsifiabilité bien moins rigide que celui que « la science » exhibe parfois en prenant appui sur une position de Popper idéalisée.

Lakatos (1970), dans le chapitre « *Falsification and the methodology of scientific research programmes* », critique la position « dogmatique » de la falsifiabilité et propose une vision plus souple de l'acceptabilité scientifique d'une théorie : « *Dogmatic falsification admit the fallibility of all scientific theories without qualification, but it retains a sort of infallible empirical basis [...] to unfalsifiable propositions the dogmatic falsificationist gives short shrift: he brands them "metaphysical" and denies them scientific standing The methodological falsificationist separates rejection and disproof [...] a theory is "scientific" (or "acceptable") if it has an "empirical basis* » (p. 172-173). L'idée n'est pas pour Lakatos de reprendre Feyerabend « *"anything goes" attitude* », il considère que nous pouvons réfuter des théories, mais prudemment, car il y a toujours des anomalies dans une théorie, parlant de « la métaphysique » de la conception d'une « *nature does not allow exceptions* ». L'objectif est d'analyser ces anomalies pour permettre au cadre théorique d'évoluer, mais pas d'avoir une position dogmatique du genre « un exemple réfute toute la théorie » (donc en opposition avec une position comme Vautier (2015g) : « *principles [...] valables pour tout système*

psychologique [...] un cas particulier peut en droit falsifier la théorie ». L'analyse des anomalies doit être vue comme une phase de « vérification » et non de « réfutation ». Pour reprendre Bourdieu (2001, p. 163) « *la connaissance scientifique est l'ensemble des propositions qui ont survécu aux objections* » (nous parlons bien « d'objection » et non d'un cadre de falsifiabilité dogmatique : c'est « vrai » ou « faux »).

Je suis donc en accord avec Vautier (2012) « *Les lois scientifiques sont des généralisations corroborées par l'expérience et il est impossible de démontrer qu'elles sont vraies. La vérité des observations, que Karl Popper (1973) appelle des propositions de base, est finalement une affaire non pas de logique mais de consensus. Les sciences empiriques reposent sur une politique de consensus.* » si on entend « consensus » au sens donné précédemment (« *qui ne peut être réfuté par personne [...] ce consensus n'a donc pas le sens d'un consensus de droit, mais de fait* », Pereira, 2012).

2.4.3.3 Pragmatisme et falsifiabilité d'un modèle

Allen & Clough (2015) discutent de la démarche pragmatique en psychologie et repositionne cette démarche comme ne pouvant pas se prétendre neutre, mais toujours construite par une pratique empirique. Ils reprennent Osbeck qui voit les scientifiques psychologiques coincés entre deux choix : accepter une théorie comme universelle ou rejeter la méthode globalement : « *Osbeck argues that psychologists too often accept a forced choice between embracing a problematic idea of method as a universal, formalist “set of rules assigned in advance”, or rejecting method altogether; the latter often in a misguided appeal to Feyerabend's Against Method (1975), understood as promoting the idea that “anything goes” (Osbeck, 2005, p. 10).* » Et de dire qu'il faut positionner la recherche en psychologie non pas dans un cadre puriste et mécaniste (« *Psychologists should not reject rationality and method per se, but instead embrace a “revisioning” of rationality understood as more humane, less mechanical, and drawing from a more diverse set of philosophical perspectives* » p. 5), mais clairement pragmatique, c'est dire positionné théoriquement et pratiquement dans un cadre global de pratiques sociales (engagement) (« *It is this practical engagement with each other and the world that gives any and all of our beliefs their meaning, their semantic content. There is no principled, substantive difference in the acquisition process by which we form beliefs concerning the empirical results of experimental manipulation, and beliefs concerning the results of philosophical or theoretical deliberation [...] For any and all of our beliefs to have*

meaning, they must have empirical content and/or they must be semantically linked to beliefs that do. », p. 9).

Il n'y a pas pour ces auteurs deux sphères séparées : les résultats vs la théorie. « *The pragmatism we reference here is more holistic than that, acknowledging a continuum of interconnected webs of beliefs, rather than splitting the two into ontologically separate spheres (the empirical/observational and the theoretical/nonobservational) that need epistemic linking* » (p. 10). Chaque théorie est chargée de croyances, mais basée sur des pratiques « *Additionally, there is no need to make the empiricist claim that beliefs about empirical observations are epistemically foundational because they are somehow theory neutral. Our pragmatist view is consistent with claims that more straightforwardly empirical beliefs are theory laden, that is, laden with/by their web-like connection to philosophical or theoretical beliefs. But the holistic view we support has theoretical beliefs laden by more straightforwardly observational beliefs as much as the reverse. And again, the model of belief webs explains the acquisition of any of our beliefs, from the observational to the theoretical, through our empirical, practical engagement with and in the world of which we are a part.* » (p. 10).

Est-ce à dire que tout est acceptable (« *anything goes* » de Feyerabend) ?

2.4.3.4 Opérationnalisation pragmatique de la falsifiabilité

Pour Pereira (2009), d'un point de vue pragmatique « *L'objectivité scientifique, comme le montre H. Putnam, est garantie par la cohérence de l'argumentation et la confrontation avec les faits. La capacité des thèses à résister à la critique argumentée, et non la neutralité axiologique, constitue alors le gage d'objectivité du savant.* » (p. 28). Nous sommes bien dans une démarche d'objectivité basée sur une « argumentation » (c'est-à-dire rationalisation) et une confrontation avec les résultats empiriques, pour questionner et « renégocier » le cadre théorique. Sans être sur une base falsificationniste dogmatique, cette démarche est falsificationniste au sens donné par Lakatos précédemment.

Nous avons donc dans la démarche pragmatique une affirmation de l'engagement de la recherche dans une pratique (contexte/enjeux), mais aussi clairement une démarche argumentée (donc contraire au dadaïsme de Feyerabend s'affirmant « contre la raison »), vérifiée empiriquement ; sachant que le regard des résultats empiriques n'est pas neutre, mais chargé du cadre du chercheur (ne serait-ce que le choix de son modèle logico-mathématique), donc à renégocier. Il y a donc une affirmation de différencier « vérité » et

« justification » comme le dit Pereira (2009) : « *H. Putnam, mais aussi Habermas, considèrent que la philosophie pragmatiste permet de maintenir la distinction entre vérité et justification sans réintroduire de fondement, ni d'absolu. Ils situent leur pragmatisme dans la lignée de celui de C.S. Peirce en faisant de la vérité la limite idéale de la justification.* » (p. 177). Ainsi pour Pereira (2009), « *Le pragmatisme est une théorie de la connaissance qui découle de la pratique : il n'y a pas de sens pour un pragmatiste à séparer théorie et pratique, pratique et justification* » (p. 141) et cite Tiercelin « *une croyance est utile si et seulement si elle est vraie et de même elle est vraie si et seulement si elle est utile* » (p. 142).

Popper a d'ailleurs discuté de « *falsifiabilité pratique* » par le recours aux méthodes statistiques (Drouet, 2012). Je ne me sers pas de cet argument pour affirmer que la démarche que je propose est poppérienne, mais pour pointer le fait que même Popper avait une vision plus souple que ce qui est parfois considéré sur la falsifiabilité. La justification empirico-pratique d'une mesure se base sur les résultats empiriques liés au cadre théorique. A l'inverse de Feyerabend tout n'est pas acceptable. En phase avec Lakatos, il me semble que nous pouvons dans un tel cadre poser des règles pratiques pour « justifier » une mesure en psychologie, testable au sens d'une réfutation potentielle argumentée de sa justification. Cette justification et réfutation sont à considérer dans la démarche pragmatique, donc pas inscrite pour une science nomothétique, mais pour une science sociale discutant de faits sociaux renégociables. On ne peut donc pas baser la critique de la mesurabilité d'une propriété mentale en disant « prouvez-le », car cela est impossible à prouver ; mais par contre une justification est opérationnelle dans une démarche pratique et discutable : hypothèse d'unicité pratique et de densité d'une propriété mentale, et vérifier empiriquement la cohérence de cette approche. Cela revient à « l'assertibilité garantie » de Peirce comme mode opératoire de falsifiabilité, dans une démarche discursive de justification nécessaire tel que Falissard et al., (2013) le posent à l'image des études qualitatives.

Ma position est donc critique d'une démarche de « lois » et donc de falsifiabilité de lois, comme par exemple chez Lacot et al. (2015) qui posent clairement que des lois doivent pouvoir être testable par des contre-exemples. Cependant, ces auteurs posent un vrai problème de la « justification » d'anomalies en psychologie souvent sans discussion, par le simple recours au « erreurs de mesure ». Les anomalies ne réfutent pas forcément une théorie, mais elles doivent la questionner.... Mais cela renvoie par ailleurs à la tension entre « individus » et « résultats moyen ».

2.4.4 Le problème de « l'individu moyen » vs « le sujet »

J'ai essayé de montrer jusqu'à présent que poser une propriété mentale comme mesurable me paraissait potentiel pour peu que l'on questionne rationnellement une telle mesure, même sur la base de modèle probabilistes. Représenter ainsi une mesure d'une propriété mentale par un modèle faisant appel aux variables aléatoires ne me paraît pas susceptible d'être invalide à priori. Cependant, si un modèle théorique nous autorise à considérer comme pouvoir représenter une propension d'une propriété mentale sous forme d'une probabilité (ou variable aléatoire), il y a une tension entre le modèle théorique et sa formalisation statistique.

Comme je l'ai exposé, fondamentalement une modélisation statistique n'est pas équivalente au modèle théorique. La modélisation statistique intègre ses propres contraintes, notamment les contraintes associées aux estimations. Dans le cadre des propriétés mentales, nous ne pouvons considérer avoir mesuré une propension chez une personne, nous ne pouvons que prétendre représenter cette propension relativement à la population de référence (l'échantillon). De plus, le modèle exhibe des mesures en ajustant « au mieux » (au sens de critères statistiques retenus, souvent la variance minimum) un modèle statistique sur un échantillon. Concrètement notre modèle donnera une validation « en moyenne » de ce modèle. Nous aurons donc toute la différence entre « un sujet moyen » et « les vrais sujets ». C'est l'objet de la discussion de cette section.

2.4.4.1 Intra-individuel vs inter-individuel

Notre discussion n'a pas différencié pour le moment l'approche inter-individus (analyse d'une propriété mentale sur un groupe de personnes) et l'approche intra-individu (l'analyse d'une propriété mentales longitudinalement pour une personne). Le cadre est identique sur le plan statistique (nous travaillons sur des variables aléatoires), mais pratiquement l'analyse sera différente car en « intra » on analyse l'évolution d'un individu. Or, la prise en compte des variabilités individuelles (Juhel & Rouxel, 2015) et la prise en compte des effets intra-individus dans le temps (Molenaar, 2004) semblent récentes en psychométrie. Comme le remarque Lamiell (2006), il y a avec les variables latentes l'illusion du paradigme « néogaltonien » qui consiste à croire que « *les résultats statistiques obtenus à partir d'agrégats lors d'études corrélationnelles des différences individuelles ouvrent une "fenêtre" épistémique sur ce qui se passe chez les individus constituant cet agrégat.* » (p. 338). La plupart des processus psychologiques ne semblent pas ergodiques (homogénéité et stationnarité des distributions de tous les individus, voir Molenaar, 2004).

Ainsi, avec un modèle reliant différentes variables latentes, on peut valider des relations inter-individus, mais pas les processus intra-individus (Borsboom, Kievit, et al., 2009). La relation structurelle trouvée entre propriétés mentales par un modèle statistique ne peut donc pas être associée au processus individuel (Borsboom, Kievit, et al., 2009; Markus & Borsboom, 2013a). Les analyses inter-individus ne peuvent donc pas prétendre expliquer les mécanismes psychologiques internes à l'échelle individuelle car toutes les propriétés mentales sont multi-réalisables et multi-déterminés (Borsboom, Kievit, et al., 2009). Comme le dit Lamiell (2006) : « *les régularités empiriques décelables par les méthodes statistiques propres à la psychologie contemporaine des traits ne constituent jamais une connaissance des personnes.* » (p. 353).

Borsboom, Kievit, et al. (2009) illustrent ce problème par le concept de « personnalité » que la dynamique individuelle de la propriété mentale telle que la psychanalyse peut le comprendre (pour ex : analyse freudienne) diffère complètement de la modélisation relationnelle de cette même propriété au niveau d'une étude inter-individu (personnalité analysée par le *big-five* par exemple). Il y a deux niveaux d'analyse, et l'analyse individuelle doit se faire dans une perspective d'interaction sociale alors que l'analyse globale détecte des différences de dispositions indépendamment de leur fonction dynamique réelle à l'échelle individuelle. Borsboom, Kievit, et al. (2009) concluent que la psychologie au niveau inter-individus et la psychologie au niveau individuel peuvent s'enrichir, mais ne peuvent converger dans leur résultats car les résultats observés au niveau inter-individus ne sont absolument pas extensibles au niveau individuel. De ce fait, la psychologie au niveau agrégé et la psychologie au niveau individuel ne peuvent que rester deux champs différents (Borsboom, Kievit, et al., 2009; Lamiell, 2006).

2.4.4.2 Conclusion

Toute cette discussion renvoie à la tension entre "l'individu moyen" et l'individu en tant que « singularité ». Falissard (2012) : « *Le sujet statistique est un sujet moyen. Sur un grand nombre de sujet, en moyenne ça fonctionne. Les irrégularités se moyennent avec le sujet moyen [...] Nous avons tendance à nous représenter [le patient] comme le sujet moyen, nous adhérons au sujet moyen ("se coller au sujet moyen") [...] Nous avons été obnubilés par le sujet moyen et non par le patient.* ». Falissard revient dans une autre conférence sur ce problème, Falissard (2014) : « *On ne montre pas qu'un médicament agit sur un patient. On montre que le médicament agit sur un sujet moyen qui n'existe pas. C'est massivement*

efficace statistiquement. Ça marche macroscopiquement. La statistique produit des portraits robots. On recherche dans le patient le portrait-robot qui lui ressemble car on a la thérapie la plus efficace statistiquement [...] La statistique montre une relation causale sur un sujet moyen qui n'existe pas [...] On utilise les mots "causalité" et "preuve" pour soutenir la pensée statistique, qui est efficace mais qui pose problème en termes de singularité. »

En accord avec cette conclusion, je ne peux que critiquer les travaux qui fréquemment postulent avoir analysé des mécanismes psychologiques individualisables à partir de modèles avec variables latentes ; alors que les résultats ne sont absolument pas transposables à l'échelle individuelle, ce ne sont que des résultats moyens, des mesures de disposition relatives à une moyenne. En quelque sorte nous ne pouvons passer de « propensions » moyennes à l'individu, celui-ci doit toujours s'appréhender comme une singularité. Cela n'interdit pas de questionner au niveau individuel de tels modèles, mais il ne peut y avoir, sur la base de résultats d'un modèle avec variables latentes, que des questionnements sur les mécanismes individuels, soit une démarche abductive qui n'est ni déductive ni inductive ; en rupture avec une démarche normative sous-jacente à toute modélisation.

2.4.5 Modèles avec variabilité intra

La psychologie développe, avec le recours de la sophistication algorithmique permise aujourd'hui, des modèles cherchant à intégrer la variabilité « intra ».

2.4.5.1 L'intégration de la variabilité intra

La psychologie différentielle a pour objet l'analyse de variance/invariance intra-individuel. Ainsi *« l'agrégation d'observations effectuées par moment chez des individus dont le comportement est "ramené" d'une manière ou d'une autre à une tendance centrale [...] ne correspond pas en lui-même au type d'hypothèse qu'on peut trouver dans une théorie scientifique explicitant des processus de développement au niveau individuel. Le niveau d'analyse approprié à cette perspective est celui de l'individu chez qui on compare les observations effectuées à des moments différents au cours du temps. L'étude des variations intra-individuelles devient ici nécessaire à l'identification d'invariants pouvant rendre compte du développement individuel dans ses aspects idiosyncrasiques. »* (Juhel, 2013a, p. 16). Seules des études longitudinales semblent permettre une telle approche.

Je m'appuie sur la synthèse de Juhel (2013a) et son exposé illustré (Juhel, 2013b) pour aborder les modèles intégrant la variabilité intra et la variabilité inter usant de variables

latentes. Je ne suis absolument pas expert des modèles discutés par Juhel. Mais j'ai une certaine connaissance des modèles stochastiques, des classes latentes, des modèles bayésiens, ... et je décèle des nœuds potentiels :

- Je rappelle qu'un modèle statistique est une écriture dans le formalisme logico-mathématique d'une théorie. Ce formalisme oblige à poser des hypothèses statistiques, notamment sur les lois des variables intégrées. Il y a donc toujours un choix « arbitraire », ce que Juhel (2013a) considère avec raison « *l'adéquation entre les hypothèses développementales et les caractéristiques des modèles statistiques permettant de les tester* ». Cela ne pose aucun problème en soi, mais plus le modèle se perfectionne, plus le nombre de choix augmente et rend cette adéquation délicate dans la pratique ;
- L'usage de modèles plus simples renvoie dans la littérature académique des effets de « conformisme » assez dramatique (Sijtsma, 2012b). La plupart des usagers de modèles avec variables latentes n'ont en effet qu'une culture faible des soubassements de la statistique mathématique de ces modèles. On arrive donc, et ceci de façon assez récurrente pour ce qui est un champ que je maîtrise relativement correctement (les SEM), à des modèles absolument erronés publiés dans des revues de référence. La complexification des modèles en intégrant des classes latentes, des cadres bayésiens, des modèles stochastiques longitudinaux, ne fait qu'accroître la distance entre la connaissance réelle des usagers et la connaissance nécessaire pour maîtriser ces modèles ;
- Ces modèles usent d'algorithmes de plus en plus sophistiqués pour estimer les paramètres. Ces algorithmes deviennent de véritables boîtes noires... et parfois susceptibles de donner des résultats erronés. De ma propre expérience, des logiciels statistiques comportent parfois des erreurs d'algorithmes que seule une démarche comparative entre différents programmes permet de révéler.
- Le problème de la sophistication des modèles statistiques pose toujours une question élémentaire : la croissance du nombre d'indicateurs estimés fragilise l'ensemble des estimations. La fiabilité statistique exigera une croissance des échantillons souvent inaccessible d'une façon pratique

La technique devenue usuelle pour estimer la fiabilité des estimations est le bootstrap, qui demande parfois un temps machine réel pour des modèles complexes. Mais le bootstrap ne révélera pas la robustesse d'un modèle, c'est-à-dire si le paramétrage utilisé permet de

considérer que les estimations sont potentiellement biaisées. La technique devenue commune pour vérifier la robustesse de modèles sont les simulations Monte Carlo (je développerai plus en détail ces méthodes dans la seconde partie de ce travail).

2.4.5.2 Revue succincte sur la robustesse des modèles

Une revue de littérature est complexe à faire sur les modèles intégrant la variabilité intra et inter, entre autres à cause de la variété des modèles potentiels (voir par exemple le spectre des modèles possibles proposés par Mplus), mais aussi surtout parce que certaines revues centrales, comme par exemple *Structural Equation Modeling: A Multidisciplinary Journal* ou *Multivariate Behavioral Research* sont devenues inaccessibles. Cependant, à partir d'un certain nombre d'abstracts récents (car je n'ai pas réussi à récupérer les articles), nous pouvons estimer la robustesse relative de certains modèles sur la base de simulations Monte Carlo. Cette revue ne prétend donc pas apporter des réponses consolidées sur ces modèles, mais permet de corroborer les problèmes usuels en statistique entre fiabilité des estimations et croissance des paramètres estimés.

Les *Latent Growth Modeling* ont fait l'objet de différentes simulations Monte Carlo qui permettent certainement de considérer pouvoir avoir aujourd'hui un cadre paramétrique précis dans lequel ces modèles sont robustes (Ke & Wang, 2015). Cependant la taille d'échantillon semble devoir être importante (1500 voire 3000) pour estimer de façon fiable certains paramètres (Diallo & Morin, 2015; Peugh & Fan, 2015). Les modèles avec classes latentes semblent aussi populaires, mais semblent susceptibles de donner des résultats biaisés (S.-Y. Kim, 2012; Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Wu & Lang, 2015) sur l'estimation du nombre de classes réelles ou autres paramètres du modèle. Les modèles markoviens (*Latent Transition Analysis*) semblent avoir fait l'objet de peu de travaux validant leur robustesse par simulations, cependant il semble acquis qu'il faille des échantillons importants, de l'ordre de 1000 au regard des quelques travaux trouvés sur la base de simulations (Cho, Cohen, Kim, & Bottge, 2010; Hyatt & Collins, 2000; Lanza & Collins, 2008). Mais ces modèles « *postulent la stabilité au cours du temps des paramètres individuels [...] ce qui en fait des modèles stationnaires au niveau intra-individuel* » (Juhel, 2013a, p. 19). Ce ne sont in fine qu'une sophistication des modèles standards avec variables latentes classiques et comme le dit Juhel : « *Les résultats obtenus grâce à l'emploi de méthodes statistiques permettant de démêler les diverses formes de variabilité, [...]*

L'interprétation théorique de la variabilité intra-individuelle est cependant parfois équivoque » car basé sur des hypothèses fortes peu vraisemblables.

Comme discuté précédemment, c'est justement cette hypothèse de stationnarité qui est contestable (non-ergodicité). D'où l'intérêt de poser des modèles permettant d'intégrer cette non-stationnarité, par exemple en intégrant des processus de type ARMA. Ces modèles ouvrent des possibilités car contournent les écueils des modèles classiques qui « *suppose[nt] que les processus développementaux sont temporellement homogènes et que le comportement de tous les individus de la population obéit aux mêmes lois dynamiques. Or cette hypothèse dite d'ergodicité des processus développementaux n'est généralement pas respectée* » (p. 25). Les approches intégrant les variabilités individuelles contournent ces problèmes : « *Ainsi conçue, l'analyse de niveau individuel des processus développementaux évite à la fois le risque d'erreur écologique ("ce qui est vrai pour la population l'est aussi pour l'individu" ; Robinson, 1950) et celui d'erreur individualiste ("ce qui est vrai pour l'individu l'est aussi pour la population"). L'objectif de l'approche de niveau individuel de la variabilité intra-individuelle est d'identifier des régularités psychologiques et comportementales plus générales à partir de l'analyse rigoureuse et cohérente des similitudes et des différences entre structures de variations intra-individuelles (Nesselroade & Molenaar, 1999 ; Nesselroade, 2010). C'est donc bien un rapprochement qui est recherché entre les niveaux normatif et individuel grâce à l'agrégation informée par "les différences et les similitudes entre patterns de changement intra-individuel" (Nesselroade, 2004, p. 228) des données individuelles* » (p. 25). Ces modèles offrent des possibilités prolixes sur les processus de séries temporelles intégrables (autorégressifs, moyennes mobiles, ARIMA, ...). Peu de simulations semblent avoir été faites. Si ne pas intégrer de tels processus semblent biaiser les résultats (Koo & Leite, 2014), intégrer ces processus n'est pas sans poser de questions sur la fiabilité des estimations (Koo & Leite, 2014; Wu & Lang, 2015). De toute façon, de tels processus exigent des données sur des séries longues pour estimer les paramètres de ces processus ; ce qui rend délicat leurs applications empiriques en psychologie.

2.4.5.3 Conclusion

Ma conclusion se voudra prudente pour ces modèles intégrant les variations intra et les variations inter. La recense de ces modèles, le peu de simulations testant leur robustesse, la complexification des modèles risquant de fragiliser les estimations, le nombre de modélisations potentielles compte tenu du nombre de modèles proposés par les logiciels

(Mplus ou R), la difficulté d’appréhender concrètement les enjeux théoriques de ces modèles relativement au paramétrage associé, me laissent considérer que si ces modèles sont des ouvertures pour l’avenir, la prudence d’utilisation me semble nécessaire. Seules des simulations Monte Carlo consolidées permettraient d’avoir un regard aiguisé sur ces modèles. Or de telles simulations ne peuvent se faire que sur un nombre restreint de paramètres pour un plan d’expérience opérationnel pour de telles simulations, ce qui pose la question de l’efficacité de telles simulations en dehors de répliques variées ; ces simulations demandent donc confrontations et discussions. Lorsque l’on étudie l’ensemble des simulations Monte Carlo sur les SEM classiques par exemple, le temps nécessaire et le nombre de simulations nécessaires pour arriver à poser un cadre considéré comme fiable de leur usage (taille d’échantillon, nombre de variables à intégrer, fiabilité des GOF, biais potentiels, ...), cela laisse deviner, qu’au regard de la croissance des paramètres en jeu dans les modèles les plus sophistiqués intégrant des variabilités intra non stationnaires, il faudra encore un certain temps avant d’avoir ce recul nécessaire pour un usage raisonné de tels modèles.

2.4.6 Echelle de mesure et tests

La psychologie utilise le terme générique de « test » pour des protocoles très différents. Or le cadre différent de ces mesures laisse subodorer que leur efficacité n’est pas forcément identique.

2.4.6.1 Test et échelle de mesure

Parfois un test est une série de mesures sur des « habilités » à réussir certaines tâches. Parfois les tests renvoient à la mesure d’une « attitude » ou d’un « trait » (les termes sont utilisés parfois de façon équivalentes dans les articles) via des « échelles de mesure » (« *scale* » en anglais). Si la littérature ne différencie pas le plus souvent ces mesures entre « test » et « échelle de mesure », fondamentalement une échelle de mesure se différencie des tests d’habileté dans l’approche de l’inférence d’une propriété mentale. Sijtsma (2011) (en lisant “questionnaire” pour ce que j’ai appelé “échelle de mesure”) « *Psychological measurement instruments are called tests or questionnaires. A test requires maximum-performance— the person is instructed to do the best (s)he can. This is relevant in intelligence measurement, cognitive ability measurement, and educational testing. A questionnaire requires typical behavior; that is, the behavior a person usually exhibits when placed in a particular*

situation— the person is instructed to show who (s)he is » (p. 1211). Cette nature différente entre “test” et “échelle de mesure” renvoie à mon avis à des problèmes différents quant à leur usage pratique. Or il semble n’y avoir que très peu de travaux discutant des qualités comparatives des tests vs échelle de mesures (disons que ma recherche bibliographique a été quasi bredouille). Donc, le développement qui va suivre mériterait d’être consolidé par une recherche plus approfondie. Je ne pose que certaines questions sur l’efficience des mesures telles que pratiquées pour amorcer une discussion qui mériterait un travail plus approfondi.

2.4.6.2 Les tests d’habilité

Un test contraint le répondant à utiliser des processus cognitifs qui seront de natures différentes pour les différentes questions d’un même test. Ce recours à des processus différents n’est pas forcément un problème en soi car c’est au niveau de l’émergence de la propriété que la discussion se pose. Cependant, la mesure d’une habilité n’est pas sans poser de problèmes.

Van der Maas et al. (2011) discutent du concept de « l’habilité » qui est la mesure qu’un test se propose de mesurer. Ces auteurs font remarquer que dans le cadre classique des variables latentes, l’habilité n’est pas une valeur absolue mais une différence entre des niveaux d’habilité (comme je l’ai exposé sur « ce que mesure un test »). Et ils posent la question : « *What is ability at the level of an individual?* » (p. 343). L’habilité est une capacité (“power”) avec un degré plus ou moins grand. « *In the terms of philosophers like Harre’ (Harre’ & Madden, 1975), at the individual level, to ascribe ability to an individual is to ascribe that individual a causal power. Although such powers may be present to a greater or lesser degree, they have a definite minimum, namely absence.* » (p. 344). A partir de ce cadre axiomatique sur l’habilité, l’unidimensionnalité d’une variable latente est questionnable pour un test pour ces auteurs : « *Arithmetic items for addition and multiplication simply cannot be unidimensional, because they depend on discretely separable abilities, each of which is plausibly governed by a separate positive ability model. The appearance of unidimensionality probably results from the fact that these abilities are strongly intertwined and arranged in a hierarchical fashion; this, in turn, results in data that appear to be unidimensional because of the implied strong positive association between item responses* » (p. 354). De nouveau émerge à mon sens la confusion entre les processus en amont et la propriété émergente objective. Pour eux, on peut néanmoins mesurer les différences d’habilité par un scalaire, donc statistiquement unidimensionnelle, sans pour autant considérer qu’une simple habilité

est mesurée. L'hypothèse est faite de construire des modèles hiérarchiques, ce qui renvoie aux systèmes dynamiques complexe déjà discuté.

2.4.6.3 Le problème de l'unicité avec des tests

Néanmoins le problème du recours à des processus cognitifs différents pose question même sur l'émergence d'une propriété. En effet, s'il apparait problématique de chercher à inférer une propriété émergente par les processus en amont (car l'émergence n'est pas réductible aux processus en amont), la question se pose de savoir si des questions activant divers processus peuvent générer des propriétés émergentes différentes.

La question centrale est de savoir si nous pouvons qualifier comme une seule et unique propriété émergente les réponses à deux tâches successives d'un test. Je me positionne donc en opposition avec Trendler (2013) par exemple (qui cite Hambleton, Swaminathan, and Rogers, 1991) pour critiquer l'unidimensionnalité de l'habilité : « *These factors might include level of motivation, test anxiety, ability to work quickly, tendency to guess when in doubt about answers, and cognitive skills in addition to the dominant one measured by the set of test items* ». Néanmoins, est-ce que la résolution de la tâche « 5 + 3 », qui peut être appréhendée comme un attracteur d'un réseau de neurones (voir Falissard, 2008a), donc comme une propriété émergente, est la conséquence de la même propriété émergente que « 10 + 2 » ? Le problème est d'autant plus compliqué que pour une seul et même tâche (« 5 + 3 »), nous savons que le processus cognitif peut être différent pour une seule et même personne : « *différentes configurations de réseaux neuronaux sont utilisés par différentes personnes pour résoudre un même problème. Selon le contexte, ou l'intention, une même personne utilise des réseaux différents pour accomplir une tâche donnée* » (Berthoz, 2013, p. 26).

Je le redis, il n'y a pas de « réalité » intrinsèque à une propriété mentale. Sa réalité est objectivée dans un cadre pratique. La question de son existence unique renvoie fondamentalement à l'indépendance locale telle que discutée antérieurement. Est-ce que les questions du test renvoient à une seule et même propriété dans le spectre des propriétés mentales théorisées au moment du test ; ou au contraire peuvent-elles renvoyer à d'autres propriétés mentales considérées comme objectivables ? Si tel est le cas, la validité du test pose question. C'est donc la granularité des habilités questionnées et appréhendées qui se pose relativement au cadre pratique. La littérature renvoie bien des critiques sur des tests usuels à travers la discussion de l'unidimensionnalité. Si la question est à mon sens mal posée

(confusion entre les processus en amont et la propriété émergente), la question est néanmoins réelle : peut-on objectiver une seule et unique propriété mentale dans un test ?

On caractérise souvent un test par un recours à « différents processus cognitifs », mais souvent ce que l'on considère être un « processus cognitif » est en fait un état stationnaire des neurones caractérisant une propriété émergente. Les processus décrits dans la littérature le sont avec des concepts et une formalisation faisant appel à des propriétés mentales et non à de véritables processus biologiques au sens d'un processus interlié de modules neuronaux. De fait, concrètement, cette discussion semble dans la littérature le plus souvent renvoyer une propriété testée relativement à d'autres propriétés potentielles auxquelles le test pourrait faire recours (et pas seulement à des processus cognitifs en amont). En effet, à hiérarchiser les items d'un test, on considère plus ou moins consciemment que le test fait appel à différentes propriétés mentales. Il y a donc à mon sens une fragilité potentielle dans les tests mesurant des habilités, car il apparaît difficile de caractériser comme propriété émergente unique les propriétés nécessaires à la résolution de l'ensemble des tâches proposées.

Cette discussion doit être menée concrètement comme je l'ai posé pour l'unidimensionnalité et l'indépendance locale, c'est-à-dire à partir de propriétés mentales objectivables au moment du test. Concrètement, sur le spectre des propriétés mentales reconnues dans notre cadre théorique, peut-on considérer que le test ne fait recours qu'à une seule de ces propriétés et ne mesure qu'une hiérarchie de cette habilité ? Ou au contraire, ne doit-on pas considérer que le test renvoie à des propriétés mentales différentes ? Et c'est justement relativement à notre approche de compétences cognitives variées que certains tests sont critiqués. Ainsi, par exemple, quand Wallot & Van Orden (2011) rappellent que pour la lecture par exemple agissent des sous composantes : sensation, perception, lecture et articulation ; on peut considérer que « sensation », « perception », « lecture » et « articulation » n'ont pas de réalité biologique stricto-sensu, mais sont bien des propriétés mentales émergentes catégorisées aujourd'hui. Un test de « lecture » donc devrait se découper en sous-test, chacun opérationnalisant une de ces sous-propriété nécessaire. Un modèle structurel avec des variables latentes de seconde ordre par exemple (construisant la « lecture » comme une méta-propriété impactant des sous-propriétés) oublierait les interactions potentielles de ces propriétés mentales de base car, comme déjà dit, la linéarité et l'additivité sont le plus souvent absentes des processus cognitifs (Wallot & Van Orden, 2011).

2.4.6.4 Échelles de mesures et le paradigme de Churchill

Une échelle de mesure semble poser moins de problème car la mesure ne s'adresse par principe qu'à une dimension unique (on mesure « ce qu'est la personne » comme le dit Sijtsma, 2011). La démarche de vérification pratique d'une mesure potentielle (que j'ai proposée antérieurement) me paraît dès lors opérationnalisable plus aisément que pour un test d'habileté.

Si l'unidimensionnalité et l'indépendance locale sont validées, le « paradigme de Churchill » (Churchill & Gilbert, 1979) est souvent l'approche de référence pour construire des échelles de mesures avec variables latentes. Ce protocole, plus qu'un vrai paradigme, propose une approche itérative entre la théorie et l'empirisme. Je résume ce protocole : l'idée générale est d'inférer un attribut latent à partir de différents items (variables manifestes) et non à partir d'un seul item. Churchill pose en premier le cadre théorique qui est de définir clairement le domaine du construit et de définir l'ensemble des items potentiels (revue de littérature, experts, études qualitatives). Cela revoie à notre discussion sur la renégociation d'une propriété mentale. Une phase quantitative sur l'ensemble des items permet de « purifier » ces derniers en ne conservant qu'une fraction d'entre eux apportant « l'information nécessaire ». Churchill discute du coefficient Alpha, mais depuis d'autres procédures (analyses factorielles avec rotations obliques, Analyse Factorielle Confirmatoire, ...) ont enrichi sa procédure. Un retour entre cette phase empirique et la théorie est nécessaire. Afin d'éviter une adaptation de l'échelle aux erreurs propres à cette première phase quantitative, une seconde phase quantitative est effectuée sur l'échelle épurée. Cette seconde phase cherche de même à tester la validité discriminante (mesure-t-on quelque chose d'original ? Le construit est-il différent d'autres ?) et la validité nomologique (le construit est-il lié à d'autres échelles de mesures, considérées comme valides, comme la théorie le prévoit ?). Ce processus de validation renvoie au cadre théorique, c'est-à-dire à notre pratique des propriétés mentales. Concrètement la validité discriminante s'effectue relativement à d'autres propriétés mentales « reconnues », donc inscrite dans le cadre pratique du moment.

Ce paradigme a été discuté initialement essentiellement dans des revues *Journal of Marketing Research* (considérée à l'époque comme la revue phare en marketing qui use des échelles de mesure en abondance) sur les outils statistiques pour valider la fiabilité de la mesure (alpha de Cronbach), voir Churchill, Peter, & Gilbert (1984); Finn & Kayande (1997); Peter (1979); Rentz (1987). Une critique fondamentale a été initiée depuis par Rossiter (2002) qui continue à être discuté par les chercheurs les plus réputés dans ce domaine : Ahuvia, Bagozzi & Batra

(2013); Diamantopoulos (2005); Finn & Kayande (2005); Francis (2009); Rigdon et al. (2011); Rossiter (2002, 2011, 2012); Rossiter & Braithwaite (2013). Cette discussion porte sur deux plans : la méthode pour valider une échelle de mesure ; le rapport entre construits mesurés réflexivement et construits mesurés formativement, point sur lequel je reviendrai dans ma discussion sur les modèles formatifs dans la seconde partie.

Rossiter (2002, 2011, 2012) apporte une critique intéressante. Le cadre usuel de construction d'une mesure est à priori : Construit → Mesure → Score. Mais de fait, le construit n'est validé que par des indicateurs statistiques (corrélations pour la validité convergente et la validité discriminante), ce qui rejoint en quelque sorte ma critique du protocole de mesure posé par Markus et Borsboom (*latent variable theory*, Markus & Borsboom, 2013a). Rossiter critique le processus concret, car le construit est ignoré en réalité dans cette démarche au sens où la théorie est ignorée : « *the construct, C, is in fact ignored, because the focus is on M and S only* » (Rossiter, 2011). Comme pour lui la validation statistique est critiquable (« *avec des critères statistiques on prend un zèbre pour un cheval* », Rossiter, 2002), Rossiter propose une nouvelle approche contradictoire avec Churchill, le C-OAR-SE, qui est formalisé sans le recours aux outils statistiques. Son paradigme pose ainsi d'autres problèmes car il basé uniquement sur des « experts », sans recours à la moindre validité empirique. Si valider une mesure uniquement par des indicateurs statistiques est critiquable, pour autant refuser de recourir à une validation empirique interdit toute confrontation avec des résultats expérimentaux.

Des auteurs, et pas des moindres (Rigdon, Preacher, Lee, Howell, Franke, Borsboom) ont écrit un article en commun pour répondre à Rossiter (Rigdon et al., 2011). Ces auteurs adhèrent consensuellement à la critique du dogme statistique posé par Rossiter : « *we applaud the critical spirit of MMR [Marketing Measurement Revolution] and at least partly agree on Rossiter's diagnosis of the pathology* » (Rigdon et al., 2011). Cependant, ces auteurs critiquent le nouveau dogme proposé par Rossiter. Pour ces auteurs, le protocole de Churchill est effectivement, comme Rossiter le critique, pris comme un dogme par bien des chercheurs, alors qu'il n'est adapté qu'à certaines mesures. Mais Rossiter propose en opposition un autre dogme aussi problématique. Comme le rappellent Rigdon et al. (2011) ce n'est pas forcément le modèle de Churchill qui est critiquable, mais la pratique « *For our part, we absolve Churchill, and place the blame squarely on the editors, reviewers, and researchers who turned Churchill's recommendations into a caricature* » (p. 1595). La mesure peut s'aider pour eux des outils statistiques, mais notre imagination est encore limitée : « *methods are available for modeling nonlinear relations, even within factor models (Muthen and Muthen,*

2010), and nonlinear logit or probit forms may make sense when either measure or construct is discrete (Reise et al., 1993) [...] researchers could acknowledge the limitations of their prior knowledge by adopting relaxed models, whether factor models with additional parameters (Asparouhov and Muthen, 2010) or the set correlation method proposed by Cohen (1982). In Asparouhov and Muthen's (2010) approach to Bayesian confirmatory factor analysis, the strictly-zero cross-loadings of conventional CFA are replaced by parameters with near-zero expected values and small variances, allowing for CFA models which do not impose a strict simple structure but which perhaps better represent actual measurement phenomena [...] a world with multiple alternative measurement models requires researchers to give serious *à priori* consideration to their choice of measurement model rather than rely on tradition and dogma » (p. 1592-1593-1594). Nous retrouvons dans ces phrases des aspects similaires à ce que j'ai développé sur les modèles statistiques et leur validation qui doit renvoyer au sens, que seule la théorie amène... mais aussi sur la complexité de poser un modèle statistique à partir d'un modèle théorique, car l'ensemble des choix statistiques rend délicat la justification théorique d'un choix vs un autre.

Mais plus fondamentalement la discussion entre Rossiter et ses détracteurs caractérise la tension de la pratique usuelle de la mesure (basée sur des validations statistiques, critiquable comme le pose Rossiter). Ce que je propose comme approche de validation d'une mesure est en rupture avec le courant dominant (l'approche statistique), mais de même en critique avec Rossiter. Il faut valider la mesure, c'est-à-dire valider une représentation potentielle par un processus rationnel de justification (et pas seulement sur la base d'experts comme le propose Rossiter). Mais au-delà de la polémique sur le paradigme de Churchill, l'approche de Churchill me paraît opérante si elle arrive en aval de cette procédure de justification de la mesure, et donc pas comme justification par elle-même d'une mesure... ce que font pourtant la plupart des articles utilisant le paradigme de Churchill.

2.4.6.5 Validité d'une échelle de mesure

La discussion sur le paradigme de Churchill et la critique de Rossiter renvoie donc à la discussion posée sur les modèles statistiques, leur validation empirique et la confusion avec le modèle théorique. L'extension des modèles, le spectre du possible (comme ceux développés par Muthen et Muthen, via MPlus) pose ce problème de façon plus prégnante car la formalisation logico-mathématique dépendra du choix du chercheur (choix du modèle, choix des paramètres, choix de critères de validation). Donc le retour vers le modèle théorique est

plus sensible car des choix différents peuvent renvoyer des résultats très différents. De plus la robustesse de certains modèles, comme déjà dit, mériterait d'être validé pour avoir un cadre paramétrique précis de son domaine de validité empirique (taille d'échantillon, sensibilité de certains paramètres, validité des GOF, ...).

Toutes ces limites permettent de juger avec critique bien des modèles de mesure développés dans la littérature, car rarement des modèles concurrents sont développés, ou des validations MTMM, seuls moyens empiriques d'apporter une certaine robustesse à une modélisation. Cependant, comme déjà dit, une échelle de mesure renvoie à la mesure d'une propriété mentale considérée stable et objectivée comme unidimensionnelle. En quelque sorte le cadre du paradigme de Churchill, et les modèles afférents, cherchent à « envelopper » la propriété via des manifestations différentes, mais liées à un domaine homogène (la propriété mentale mesurée). Comme justement cette propriété mentale ne se révèle que par ses manifestations, il est assez aisé d'essayer de l'inférer par ses manifestations. Bien évidemment cette enveloppe de manifestations dépend de la caractérisation potentielle de la propriété mentale et de sa mesure (nominale, ordinale, continue) ; des variabilités potentielles intra ; du cadre hypothétique (modèle bayésien par exemple). Mais une échelle de mesure présuppose une forme de stabilité chez la personne, ou du moins la capacité de caractériser une propriété manifestée comme telle. Sans remettre en cause la difficulté et bien des critiques de telles mesures, il m'apparaît potentiel de pouvoir prétendre effectuer des mesures que je considérerai comme « sereine » par une échelle de mesures, dans le sens où la mesure renverra une représentation fiable de la propriété mentale objectivée.

2.4.6.6 Test et échelle de mesure

Il y a donc une différence fondamentale entre test et échelle de mesure. Un test cherche à mesurer des variations d'habilité, et donc pose la question de l'unidimensionnalité de l'objet mesuré. Une échelle de mesure renvoie de façon plus évidente à l'unidimensionnalité car il n'y a qu'une seule propriété mesurée. Cette différence explique à mon sens des résultats contradictoires rapportés par la littérature.

Van der Maas et al. (2011) développent un des rares travaux à ma connaissance comparant la mesure d'une habilité vs une échelle de mesure (attitude). Pour mesurer l'habilité, pour ces auteurs il faut que l'individu ait cette habilité à un certain degré (qui peut être nulle), cela diffère des mesures d'attitude pour eux car, d'une façon axiomatique, l'habilité vérifie : 1/ l'habilité est présente ou absente ; 2/ toute tâche pour mesurer une habilité nécessite cette

habilité ; 3/ une tâche requérant une habilité détenue par une personne peut être effectuée par la personne si elle a assez de temps pour ce faire. La mesure d'une attitude présuppose que cette attitude existe, mais de façon plus ou moins prononcée (il n'y a pas nullité potentielle de l'attitude) ; et que la mesure révèle cette attitude indépendamment d'un temps d'exécution (à la différence des tests). Ces auteurs ont recours à un modèle (« Q-diffusion ») qui intègre le cadre axiomatique du test et de sa mesure. Leurs résultats montrent qu'un test semble non-fonctionnel avec le modèle IRT, mais fonctionnel pour des mesures d'attitude : « *This is surprising because IRT models have been traditionally proposed for, and applied in the context of, ability testing* » (p. 353).

Ce résultat mériterait d'être confirmé par d'autres travaux. Mais il y a en effet un facteur de risque plus important dans la mesure d'un test, car l'unidimensionnalité, et l'indépendance locale inhérente, est bien plus problématique à poser. Une échelle de mesure peut répondre, comme déjà dit, au cadre actualisé d'une mesure (unidimensionnalité/indépendance locale). Si nous sommes en droit de considérer, inscrit dans une pratique clairement posée, l'objectivité d'une échelle de mesure unidimensionnelle, la mesure me paraît bien plus aisée. Juhel & Gilles (2011) rappellent que les items d'un test : « *Par ailleurs, la représentativité des items qui composent l'outil fait référence au degré avec lequel ceux-ci échantillonnent l'univers des items possibles par rapport au domaine du construit mesuré (Dickes et al., 1994). Ceci implique la disponibilité d'une définition d'un univers de contenu et d'items ainsi que d'une procédure d'échantillonnage (ou des critères de sélection) des items, de sorte que ceux constituant l'outil soient représentatifs de l'univers d'items ou des dimensions ou catégories qui structurent cet univers (le plus souvent, des sources de variations interindividuelles)* » (p. 12). Et c'est bien la représentativité de l'univers qui peut être questionnée pour certains tests d'habilité car s'il y a différentes « catégories », nous ne sommes plus sur une propriété mentale unidimensionnelle.

En conclusion, autant une échelle de mesure posée comme je l'ai proposé (validation empirique d'un homomorphisme entre la propriété et une échelle numérique) me paraît opérationnalisable ; autant la mesure d'habilité nécessite une discussion bien plus approfondie sur le protocole utilisé. Cela n'interdit pas à mon sens des tests d'habilité, mais avec une renégociation permanente des tests et des items en relation avec le spectre des propriétés potentielles.

2.4.7 Conclusion

Je défends la possibilité de mesurer des propriétés mentales dans une démarche clairement pragmatique au sens du courant de la philosophie des sciences et non du pragmatisme au sens commun. Ma position se veut donc critique des travaux considérant que mesurer est par essence en psychologie une « pathologie » (Michell), une « hérésie » (Maraun) ou un « fétichisme » (Vautier). Mais cette possibilité de mesurer n'implique pas que toute mesure se justifie. De même, la possibilité de mesurer n'implique pas que les résultats des modèles statistiques opérationnalisant une mesure soient fiables par essence. Ainsi, la critique de la mesure portée par certains auteurs renvoie à de véritables problèmes de la pratique usuelle de la mesure.

2.4.7.1 La démarche de mesure

Je considère que nous pouvons mesurer une propriété mentale si elle est catégorisée comme unidimensionnelle. Comme expliqué, cette catégorisation est inscrite dans une pratique et n'est pas vraie « en soi ». C'est la proximité relative de manifestations qui nous fait considérer comme propriété unique sous-jacente l'origine de telles manifestations. Mais demain, une renégociation de ces manifestations nous fera peut-être catégoriser et conceptualiser ces manifestations comme relevant de différentes propriétés mentales. Il n'y a pas de « vrai » ou de « faux », juste une négociation inscrite dans un contexte. La réification en un objet unique est certes un artifice, mais nécessaire pour l'homme pour rendre intelligible son environnement. Cette réification s'appuie sur une réalité, les manifestations, et par un jeu de boucle entraîne cette réification à devenir une réalité (Hacking, 2008). Ce n'est donc pas une démarche constructiviste, mais bien réaliste... avec un petit « r ».

A partir d'une telle propriété mentale catégorisée, la question d'une continuité empirique, c'est à dire d'une densité ou non des différents états potentiels, permet de poser une mesure (nominale, ordinale ou numérique) en tant que représentation « fidèle » (au sens de notre appréhension de cette propriété au moment de sa réification). La mesure s'effectue par un modèle statistique, basé sur des variables aléatoires caractérisant des « propensions », donc en opposition avec des modèles déterministes. La validation et la falsifiabilité s'appuient sur des résultats empiriques reliés, par une démarche réflexive, au cadre théorique.

2.4.7.2 Divergence avec la pratique usuelle

La divergence, entre ma proposition et ce que la littérature discute le plus souvent, provient fondamentalement de la nature ontologique des propriétés mentales. L'empirisme-réaliste pose de façon plus ou moins explicite que la propriété mentale est une « réalité à découvrir », pour exemple Juhel & Gilles (2011) caractérisent qu'un test mesure « *des entités réelles, en attente d'être découvertes* ». Or, comme je l'ai discuté, il n'y a pas d'entité « réelle en attente d'être découvertes » en psychologie au sens d'objets fixes inobservables qui seraient à découvrir, comme la physique le fait par exemple avec les électrons ou particules élémentaires. Les propriétés mentales sont des propriétés émergentes inscrites dans une praxis sociale et donc à négocier et renégocier.

De cette confusion sur la nature de « réalité » de ces propriétés mentales découlent que l'ensemble de la littérature révèle la confusion entre les processus cognitifs en amont, processus complexes qui ne peuvent pas se mesurer, et la propriété mentale émergente, qui se manifeste en tant qu'entité irréductible aux processus en amont, et qui elle peut être mesurable. C'est cette caractérisation des propriétés mentales comme entités émergentes qui renvoi l'efficience des modèles avec variables latentes.

Mais ce que je propose comme cadre pratique peut in fine ne pas paraître très éloigné de la pratique telle que proposée par exemple par Markus & Borsboom (2013a) et couramment utilisée. Mais la première différence, et cela en phase avec toute la critique portée par Michell par exemple, c'est l'obligation de justifier la mesure utilisée. Or le livre de Markus et Borsboom ne discute pas de la démarche empirique concrète pour valider une mesure quantitative. De même, dans leur rapport sur la validité des modèles psychométriques, Juhel & Gilles (2011) ne discutent pas de la manière de justifier la quantification d'une propriété mentale. Il y a une sorte d'admissibilité tacite dans la littérature que toute propriété mentale peut se représenter sur un continuum numérique. Or rien ne permet de l'affirmer. Il faut discuter pour une propriété mentale ordinale d'une densité empirique potentielle pour pouvoir la considérer comme quantitative.

Cette non-justification d'une mesure quantitative n'est peut-être pas déconnectée de ma seconde différence avec la démarche usuelle. Je considère que la mesure d'une propriété mentale doit s'effectuer dans la renégociation de cette propriété. Or la plupart des articles usant de tests ou d'échelle de mesure utilisent des mesures « validées » antérieurement, comme si une propriété mentale pouvait être posée sans exiger sa renégociation, comme si elle était indépendante d'une pratique inscrite dans un contexte.

Enfin, mais je ne fais que reprendre ce que toute une littérature expose, il faut clairement différencier le modèle statistique, le modèle théorique et la réalité. Il y a une discussion critique nécessaire à avoir sur les choix du modèle, il faut discuter de modèles concurrents (ce qui est rarement effectué), discuter d'une réplication de la mesure (MTMM), qui fait là aussi souvent défaut. De même il faut un regard critique, que la littérature pointe depuis quelques temps, sur la confusion entre « l'individu moyen » et la « singularité ».

2.5 Critique de l'usage social de la mesure

Si la mesure d'une propriété mentale est inscrit dans une pratique ; l'usage généralisé des mesures en psychologie doit donc être abordé comme une pratique sociale. C'est-à-dire relativement au contexte et enjeux de cette pratique à l'échelle de l'espace public.

2.5.1 Impératif quantitatif ?

La science, qui prend comme parangon la physique, est souvent appréhendée comme devant baser ses raisonnements sur une représentation logico-mathématique de la réalité. La mesure en psychologie est fondamentalement inscrite dans cette vision de la production scientifique.

2.5.1.1 Langage logico-mathématique et critique de l'impératif quantitatif

La fonction des mathématiques est de construire un discours épuré d'un contexte socio-culturel : « *mathematics basically examines idealized objects (most often numbers and shapes), which can be structured and related one to the other. It is notable that the elements of natural language (words) can also be structured (into sentences, etc.) and related one to another (either via syntax or via a paradigm, as in structural linguistics [17]). But words are not idealized objects; they depend on both culture and physical reality [18] as well as on history.* » (Falissard et al., 2013, p. 4). Cette exigence de mathématisation renvoie à une exigence de « mesurer » en psychologie pour poser un cadre de pratique permettant le recours à ce langage logico-mathématique.

Mais au-delà, de telles mesures ont une fonction sociale : synthétiser pour contourner la complexité, résumer et classer les individus (McGrath, 2005a). Ces mesures ont des fonctions simplificatrices pour éviter la surcharge cognitive, mais aussi pour objectiver. Desrosières (2008a), dans ce sens, pose : « *La statistique était présentée comme un outils de la rationalisation [...] en substituant la raison de la mesure et du calcul à l'arbitraire des passions [...] la statistique a donc été investie d'un rôle comparable de désidéologisation et d'objectivation, permettant de traiter les faits sociaux "comme des choses" » (p. 22) et « La statistique [...] implique la mise en forme de schèmes scientifiques (moyenne, dispersion, corrélation, échantillonnage probabiliste), destinés à résumer, notamment par des outils mathématiques, une diversité supposée non maîtrisable » (p. 79).*

Pour O. Martin, la quantification devient le « piège de l'évidence » : « *Ce piège de l'évidence a une double conséquence : non seulement les psychologues d'aujourd'hui ne s'interrogent pas (ou peu) sur la légitimité de la quantification, mais les historiens, voire les philosophes des sciences eux-mêmes, sont parfois tentés de prendre l'avènement de la quantification comme un phénomène naturel sur lequel il est inutile de s'arrêter* » (O. Martin, 1997, p. 2). Cependant bien des auteurs critiquent cet impératif quantitatif (pour exemple : P. Barrett (2005); J. Martin (2003); Michell (1999, 2003a, 2003b, 2005)). Michell est certainement, en psychologie, l'auteur le plus prolix sur la critique de ce présupposé qui veut que la réalité soit fondamentalement quantifiable (approche pythagoricienne de la réalité), et donc critique d'une démarche scientifique qui considère devoir se rapprocher d'une démarche mathématiques (Michell, 1999). Pour Michell, la psychologie est donc victime de cet impératif quantitatif en sciences. P. Barrett (2005), critique lui aussi de cet usage sans modération des modèles quantitatifs, rappelle par ailleurs que les grandes théories en psychologie ne sont pas nées de modèles statistiques.

2.5.1.2 Mesure et sens de la mesure

Le besoin de mesures se justifie par le recours à des modèles logico-mathématique qui semblent épurés de toute subjectivité ; les mesures utilisées en psychologie sont ainsi souvent présentées comme des « vérités scientifiques » ; les résultats statistiques des modèles sont exhiber comme des « preuves » de la vérité de théories et mesures. Or, comme dit auparavant, la pratique concrète est exactement inverse : les modèles statistiques sont tout autant inscrits dans une subjectivité que les études qualitatives. Il y a fondamentalement dans l'usage des modèles statistiques l'illusion d'avoir construit des vrais modèles explicatifs, des vrais modèles « scientifiques », à cause du recours au langage logico-mathématique. On oublie souvent que cette représentation mathématique n'est ni déterministe en psychologie, ni dépourvue d'arbitraire du fait des choix inhérents à toute modélisation statistique.

Il y a donc un leurre usuel en sciences sociales qui est de croire que la réalité est plus objective avec des chiffres. Les modèles statistiques, les formalisations par des variables, aident à la compréhension du réel, c'est une forme de représentation parfois utile. Mais les modèles statistiques en psychologie ne sont qu'un moyen d'objectiver la réalité ; de donner, par une représentation, du sens à la réalité. Cette représentation de la réalité, son objectivation par un modèle, est, comme toute catégorisation, inscrite en tension entre réalité et subjectivité. Il faut donc conserver une certaine modestie car nos modèles statistiques ne sont que des

« objectivités pour nous ». La catégorisation d'une propriété mentale est inscrite dans un cadre social ; de même, une mesure et un modèle s'inscrivent dans une démarche et une théorie qui dépendent intrinsèquement de son contexte. Il faut donc dépasser cette tension entre « quali » et « quanti » en psychologie, tel que le texte de Falissard et al. (2013) le pose : « *Psychiatry is a discipline that is particularly sensitive to this duality of words and numbers, qualities and quantities, “soft” and “hard” approaches* » (p. 1).

2.5.1.3 Mesure et quantification

Plus fondamentalement, cet impératif quantitatif et ce recours généralisé aux mesures, entraînent un travers : c'est la mesure qui crée l'objet ; on est alors dans un processus de quantification et non de mesure. Comme le dit Desrosières, il y a la confusion commune en science sociale entre « *quantification* » et « *mesure* » : « *Il faut en effet distinguer deux idées, trop souvent confondues, celle de quantification et celle de mesure. Le verbe quantifier [...] : exprimer et faire exister sous forme numérique ce qui auparavant était exprimé par des mots et non par des nombres. En revanche, l'idée de mesure, inspirée de l'épistémologie traditionnelle des sciences de la nature, implique que quelque chose existe, sous une forme déjà mesurable selon une métrologie réaliste* ». (Desrosières, 2008a, p. 10).

Or cette distinction est souvent confuse en psychologie. Bien des travaux génèrent des « valeurs numériques » par des tests croyant « mesurer » alors qu'ils ne font que construire des « indicateurs » qui ne peuvent absolument pas prétendre être des « représentations/objectivations » d'une réalité. On se retrouve donc avec une discussion dans des articles basés sur une épistémologie affirmée réaliste, mais avec des outils méthodologiques basés sur une épistémologie constructiviste. Comme le dit Desrosières : « *Un spectre hante le travail du statisticien, celui de la réalité des objets dont il exhibe la mesure* » (Desrosières, 2008a), ou Gould (1997) : « *L'essentiel de la fascination des statistiques tient dans ce sentiment viscéral [...] que les mesures abstraites résumant de grands tableaux de données doivent exprimer quelque chose de plus réel et de plus fondamental que les données elles-mêmes* » (p. 278), ou Armatte (2010, p. 20) : « *Les choses existent avant d'être mesurées et si elles n'existent pas on s'occupera de les faire exister par des conventions sociales avant de construire leur dispositif de mesure* »

Ainsi, toute quantification n'est pas une mesure, au sens de représentation d'une propriété réelle ; la quantification est une écriture numérique, mais avec des objets « construits » et non « réels » (même si des objets construits peuvent devenir en boucle des réalités de par leur

fonction). Or, la démarche concrète en psychologie semble souvent user de telles quantifications et non de réelles mesures. Au lieu de partir d'un cadre théorique inscrit dans un contexte et des enjeux, au lieu de justifier dans ce cadre une mesure (propriété unidimensionnelle ? Indépendance locale des manifestations ? Propriétés mentale représentable sur un continuum numérique ?), la psychologie construit des « indicateurs quantitatifs » qui deviennent la base de la catégorisation de la propriété mentale et de la théorie. Parlant de la mesure de l'intelligence, Juhel (2005) dit ainsi : « *L'intelligence mesurée s'est ainsi trouvée enfermée dans une tautologie dans laquelle les variations de performance aux tests sont devenues ce par quoi était désormais définie conventionnellement l'intelligence humaine* ». De même le test de Rorschach « *a la particularité d'avoir été construit avant de savoir ce qu'il allait mesurer* » (Juhel & Gilles, 2011, p. 22). Ces deux exemples illustrent la dérive fréquente où semble s'enliser la psychométrie : il faut pour bien des chercheurs impérativement des mesures pour objectiver la réalité, mais au lieu de « mesurer », les psychologues construisent des « quantités » qu'ils prennent pour des « mesures » objectivant à leurs yeux une réalité... qu'ils n'ont pas validée en amont de leur quantification.

Et donc la critique de l'impératif quantitatif de Michell semble pertinente. Avec de telles pratiques de quantification (et non de mesure), il y a une rupture avec l'hypothèse réaliste et nous sommes clairement dans une démarche constructiviste. Cela n'interdit pas de quantifier, mais en différenciant bien « la quantification » comme outil technique constructiviste, et la « mesure » comme représentation d'une réalité.

2.5.1.4 Une dérive actuelle

L'impératif quantitatif est porté par bien des chercheurs depuis longtemps : par Cattell (en 1893) : « *The history of science is the history of measurement* » (cité par Engelhard, 2008) ; Brunshvicg (1922) : « *Cette condition de mesure est préalable à toute conception, à tout langage, même d'ordre scientifique* » (cité par Allouche, 2014) ; ou encore Lord Kelvin (en 1883) : « *Je dis souvent que si vous pouvez mesurer ce dont vous parlez et l'exprimer par un nombre, vous savez quelque chose de votre sujet, mais si vous ne pouvez pas le mesurer, si vous ne pouvez pas l'exprimer en nombre, vos connaissances sont d'une pauvre espèce et bien peu satisfaisantes.* » (cité par Allouche, 2014).

Mais ce regard sur la science a été accentué par le développement informatique permettant des calculs nombreux et complexes. Pour exemple, je citerai Boyd et Crawford qui glorifient sans nuance la quantification à outrance du monde social « *c'est un monde dans lequel des*

quantités massives de données et les mathématiques appliquées remplacent tous les autres outils qui pourraient être utilisés. Exit toutes les théories sur le comportement humain, de la linguistique à la sociologie. Oubliez la taxinomie, l'ontologie et la psychologie. Qui peut savoir pourquoi les gens font ce qu'ils font ? Le fait est qu'ils le font, et que nous pouvons le tracer et mesurer avec une fidélité sans précédent. Si l'on a assez de données, les chiffres parlent d'eux-mêmes » (Boyd & Crawford, 2011).

Cette inflation de possibilités algorithmiques permet une inflation de modèles. On peut appliquer à la psychologie ce que dit Armatte de l'économie quantitative : « *En permettant l'enfermement dans un jeu purement logique de raisonnement hypothético-déductifs pouvant se poursuivre indéfiniment, les modélisations mathématiques, qui n'épuiserait jamais la combinatoire des spécifications formelles des hypothèses et des fermetures du système étudié, fabriquaient cette situation de "autiste" de la discipline* » (Armatte, 2010, p. 5). Il est d'ailleurs intéressant de voir la prudence de certains scientifiques face à l'émergence de la modélisation en sciences sociales (dans les années 1930) : « *Il est de l'essence d'un modèle que l'on n'y fasse pas figurer les variables avec leurs vraies valeurs. Si l'on procédait ainsi, on rendrait le modèle inutile. Car dès qu'on fait cela, le modèle perd sa généralité et sa valeur en tant que mode de pensée* » (Keynes, 1938 ; cité par Armatte, 2010, p. 23).

Il y a donc actuellement une dérive basée sur une vision de la quantification comme outil central de la réflexion scientifique permis par le développement informatique.

2.5.1.5 Regard critique des résultats statistiques

Il y a par ailleurs la force symbolique de la statistique contre laquelle aucune critique ne semble possible, car une valeur apparaît « exacte », sans « interprétation subjective », « c'est un fait »... oubliant que cette mesure renvoie à une construction, des choix, en amont. Desrosières (2010) développe dans son ouvrage la tension entre la genèse des outils statistiques, les polémiques associées, et la pratique actuelle : « [les] *mesures statistiques* [...]. *En tant que références, ces objets doivent être perçus comme indiscutables, au-dessus de la mêlée* ». (p. 7) et « *la statistique* [...] *est devenue synonyme d'instrument de preuve, et de référence peu discutable.* » (p. 396), ou Desrosières (2008a) : « *La quantification, signe d'objectivité, de rigueur, d'impartialité* [...] *outil de "preuve"* » (p. 7)

Ainsi, la plupart des travaux exhibe des indicateurs statistiques (corrélations, *p-values*, ...) comme des éléments de vérité, considérant ces derniers comme s'ils n'étaient pas susceptibles d'être critiquables. Une mesure est souvent présentée comme un fait inattaquable. Je peux

repandre à mon compte la critique du « *stiscism* » portée par Lamiell (2013) qui critique cette validation statistique outrancière des articles en psychologie. Or, la mesure, tout comme tous les indicateurs statistiques, dépendent de choix, d'heuristiques. Falissard et al. (2013) posent avec raison : « *The taxonomy introduced in this paper shows that the choice to use mathematics or not is fundamentally independent from the degree of neutrality of the observer: mathematics is not one but many different means that can be used to describe the world.* ».

Le discours littéraire offre bien plus de prise à la critique que le discours statistique²⁷. Je considère que l'on peut user de modèles quantitatifs, mais avec en permanence un regard critique sur ces modèles et résultats. Critique de la généralité vs la singularité ; critique de l'inertie d'un modèle vs l'irrégularité nécessaire à la vie sociale humaine ; critique de l'analyse statistique vs la faible puissance des résultats statistiques.

Ce regard critique doit à mes yeux s'affirmer de façon plus saillante que pour les études qualitatives. En effet, implicitement (ou explicitement) nous aurons un regard critique sur une étude qualitative car le biais du chercheur paraît évident. La force symbolique des chiffres et du langage mathématique, nous fait souvent oublier que nous devons avoir le même regard critique envers des modèles statistiques. Desrosières (2010) dit avec raison : « *Le retournement mental qu'implique l'attention aux métamorphoses de l'argument statistique est presque aussi difficile pour le chercheur que pour le citoyen* » (p. 396). Or aucun modèle, même ceux intégrant la variabilité intra (par des processus ARMA par exemple) n'a la capacité de décrire dans toutes ses dimensions « la réalité » et de prévoir l'avenir. Quelle que soit leur sophistication, ils ne pourront être qu'une aide à la réflexion.

2.5.2 Modèle et singularité

J'ai déjà abordé la tension entre le résultat moyen et son extrapolation à l'individu. Mais je vais développer de façon plus fondamentale les limites de la modélisation pour la singularité.

²⁷ Par exemple Gould (1997) explique que la critique du livre « *The bell curve* » (Herrnstein & Murray, 1994) par différentes personnes ne discutait pas des résultats statistiques car ceux-ci étaient « trop sophistiqués » pour la plupart des personnes qui avaient porté la critique. Or ce livre n'utilise que de régression multiple et de tests usuels, soit de modèles assez élémentaires, enseignés en niveau M1 en psychologie par exemple. Cela veut dire que la critique d'un discours statistique basé sur un modèle structurel intégrant des processus ARMA pour estimer les variabilités intra vs inter, ou des modèles bayésiens, ne pourront être critiqués que par une minorité « d'experts ». Cela limite le spectre des critiques car bien des « experts » justement sont adeptes de la modélisation statistique sans critique ; et souvent les critiques de la statistique n'ont pas la maîtrise technique de ces outils.

2.5.2.1 Invariants et singularités

Dans les mesures effectuées en psychologie, il y a souvent un processus d'indifférenciation des personnes, leur intégration à une globalité (Lamiell, 2013). Danziger (1985) alertait déjà de cette tendance à vouloir agréger toutes les personnes, à vouloir des théories générales pour les individus. Cette généralisation n'était valable pour lui que s'il y avait isomorphisme entre les individus (ergodicité en fait) et si la théorie globale avait un sens au niveau individuel. J'ai rappelé l'impossibilité statistique de passer d'un résultat moyen applicable à chaque personne à cause de la non-ergodicité des données. Les modèles intégrant la variabilité intra cherchent des « invariants » statistiques (ou du moins des formes de régularités) au niveau de la personne. De Ribaupierre (2015) rappelle que souvent, nous analysons le résultat moyen de l'individu, sans intégrer réellement les irrégularités. De plus, il est difficile de modéliser ce qui différencie « variabilité » et « changement » (Thiébaud, 2015), et il faudrait distinguer dans la variabilité « *ce qui revient à la flexibilité, la plasticité, à la vicariance ou au bruit* » (Thiébaud, 2015).

Je le redis, je considère qu'il y a du « sens » à apporter aux propriétés mentales, mais sans prétendre déterminer des lois au sens des sciences de la nature. Si la psychologie ne peut déterminer des lois déterministes, ce n'est pas seulement à cause de la complexité des processus sociaux et psychologiques, ce n'est pas seulement que les objets de son étude sont inscrits dans une pratique et donc devant être renégociés. C'est que toute modélisation statistique ne peut prédire le « nouveau » car elle ne peut que modéliser à partir d'invariances ou régularités (au sens statistique) du passé. Or l'histoire du monde social montre que c'est la création/nouveauté qui est au cœur du processus humain, donc psychologique.

Il n'y a donc pas que la tension entre un cadre généralisable (le modèle statistique) et son application sur des singularités. Il y a fondamentalement la tension entre le fait de modéliser à partir de régularités et l'émergence de la nouveauté. La psychologie ne peut donc se contenter que de proposer des relations « probables » entre certaines propriétés mentales, c'est-à-dire donner une forme de sens ; non comme cadre déterministe, mais comme cadre analytique, que ce soit au niveau « moyen » ou au niveau « singulier », pour agir.

2.5.2.2 La dimension intersubjective de la mesure

La mesure renvoie souvent au fantasme d'avoir mesuré une réalité « biologique » à l'intérieur de la personne. Or, « *Le sujet biologique n'est pas un sujet pensant* » (Falissard, 2012). La mesure en psychologie renvoie à une propriété mentale inscrite dans l'intersubjectivité. Je

renvoie à Falissard (2008b) qui, par la dénomination de « mesure subjective », permet de poser cette différence entre « subjectivité » et « objectivité » d'une mesure : « *Par définition (dictionnaire le Robert), est subjectif ce qui : "concerne le sujet en tant qu'être conscient". La subjectivité (relative au sujet pensant) est ainsi opposée à l'objectivité (relative à l'objet pensé). Voilà pourquoi, quand nous parlerons de mesures de qualité de vie, de satisfaction ou de dépression nous parlerons désormais de "mesures subjectives". Ces dernières abordent en effet le sujet en tant que sujet pensant alors que les mesures plus classiques, comme la tension artérielle, la numération globulaire ou le diamètre bitrochantérien sont relatives au sujet considéré comme un sujet vivant.* ».

2.5.2.3 Critique du portrait robot

Si la psychologie est un moyen d'intervenir dans l'interaction sociale pour appuyer des démarches personnelles (pathologie psychiatrique, aide scolaire, ...), le psychologue doit clairement différencier l'individu pensant en tant que singularité et la norme (les résultats moyens auxquels se réfèrent les mesures). Or, c'est ce décalage entre norme et singularité qui est difficile à appréhender. Les mesures sont tellement considérées comme des vérités (car basées sur des modèles soi-disant « exacts »), que le psychologue n'est souvent pas armé pour critiquer de telles mesures. Mais plus fondamentalement, la réduction de l'énergie cognitive nous pousse à user d'un cadre normatif, à oublier la singularité pour nous réfugier dans des portraits robots : « *La statistique produit des portraits robots. On recherche dans le patient le portrait-robot qui lui ressemble car on a la thérapie la plus efficace statistiquement [...] Les êtres humains aiment le consensus [...] la dictature du "on". On fonctionne comme un sujet moyen [...] "tout le monde est autrui, aucun n'est soi-même" [...] La statistique montre une relation causale sur un sujet moyen qui n'existe pas [...] On utilise les mots "causalité" et "preuve" pour soutenir la pensée statistique, qui est efficace mais qui pose problème en termes de singularité.* » (Falissard, 2014).

2.5.2.4 Critique de la norme

Ce rapport « norme statistique » vs « singularité » impose une critique de la norme en tant que construction. Il faut affirmer que les tests et modèles statistiques ne peuvent servir que de repères potentiels, apprendre aux psychologues à avoir une distanciation avec ces faits qui sont, tout comme toute démarche qualitative, des formalisation/représentation inscrites dans une pratique ; et non des « vérités scientifiques ».

Mais, plus fondamentalement, ce rapport norme/singularité renvoie à un cadre social qui très souvent a une dynamique de normalisation, anesthésiant la différence par le poids de la référence normative, qui apparaît comme la référence obligatoire. « *C'est d'abord une intériorisation consentie ou inconsciente à l'égard du principe de la mesure qui s'institue, emblématique dans le cadre du travail, le Quantified Self, ou d'autres registres du quotidien, qui conduit à discipliner de soi-même ses gestes afin de répondre au mieux aux différents critères et normes explicitement ou tacitement définis* » (Sadin, 2015, p. 116)

Il y a ce carcan à dénoncer, qui entraîne l'interdiction normative de l'émergence de la singularité. Toute cette discussion n'est pas nouvelle comme le rappelle Desrosières (2010) puisqu'elle peut être remontée à la controverse entre « nominalistes » et « réalistes » au Moyen-Âge ; ou au débat sur « l'homme moyen » au XIX^{ème} (est-il une réalité ?) et le libre arbitre vs le déterminisme (je renvoie au chapitre « *Moyenne et réalisme des agrégats* » de ce livre très instructif) ; ou au débat entre Linné et Buffon (« *l'opposition entre Linné, pour qui "toute la nature peut entrer dans une taxinomie", et Buffon, pour qui "elle est trop diverse et trop riche pour s'ajuster à un cadre aussi rigide"* », Desrosières, 2010, p. 293) ; ou au débat sur l'homogénéité vs hétérogénéité de la nature (voir de même dans le même livre de Desrosières, chapitre 9, le débat entre Keynes et Tinbergen à ce sujet ; même si Tinbergen avait visiblement conscience du « bricolage » de toute modélisation statistique).

L'idée n'est pas de refuser une vision normative comme référence (cadre « probable » ou de « vraisemblance /*likelihood* » au sens statistique) pour proposer une action individualisée, car le refus de la généralisation interdit souvent de donner du sens à de l'information, or on est en droit de rechercher ce sens au-delà des contingences individuelles. Par contre, l'idée est de refuser l'impératif de la norme comme seul schème possible de réponse à toute singularité, de refuser de ne voir, dans les anomalies entre la singularité et la norme, qu'une « erreur de mesure ». Varela (1998) écrit à juste titre : « [on ne peut] *pas comprendre la cognition si on l'abstrait de l'organisme inséré dans une situation particulière avec une configuration particulière, c'est-à-dire dans des conditions écologiquement situées* ».

2.5.3 Rôle des mesures

La critique de la mesure dans son usage social renvoie à sa fonction politique tel que dit par Desrosières (2010) : « *Il y a le langage de la description et de la science (auquel se réfère la modélisation statistique) et le langage de l'action et de la prescription qui lui est politique se servant du discours scientifique.* » (p. 413)

2.5.3.1 Enjeux de la mesure

Juhel & Gilles (2011) ont produit un rapport de synthèse sur la validité des mesures en psychologie (intitulé *Validité des modèles et des outils de l'examen psychologique*). Ce travail laisse considérer que les tests sont utilisables avec précaution, mais surtout il permet de considérer les enjeux de leur construction et usage : *« les modèles théoriques et les mesures qu'utilise le psychologue, bien qu'ils puissent être intrinsèquement valides, au sens où le terme a été précédemment défini, ne garantissent pas pour autant l'entière cohérence et pertinence des interprétations qui seront construites et des décisions auxquelles elles conduiront. En effet, les compétences du psychologue et les diverses contraintes auxquelles il a à faire face dans le contexte particulier de l'examen psychologique de l'enfant, l'amènent nécessairement à devoir faire des choix théoriques et techniques qui limitent inévitablement le champ des interprétations possibles de la situation problème. De fait, dans la pratique concrète de l'examen psychologique, la validité n'est sans doute jamais définitivement acquise : elle est un idéal vers lequel le psychologue peut (doit) tendre, un questionnement systématique sur les choix opérés et sur les interprétations qui en découlent devant y contribuer. On voit bien dès lors les enjeux déontologiques, sociaux et humains de cette question puisque, au final, ces choix et ces interprétations ont des conséquences directes sur la nature des décisions que le psychologue aura à prendre au terme du bilan psychologique. »* (p. 34-35). On retrouve ici posée la tension des choix du psychologue.

Juhel et Gilles rapportent que pour les psychologues ces mesures sont « utiles ». La question est la raison de cette utilité. Comme le soulignent les auteurs de ce rapport, la plupart des tests orientent des actions. Et donc, en phase avec Désrosières, l'action renvoie à une dimension politique de l'usage de la mesure et donc de son « utilité », relativement à l'action envisagée.

2.5.3.2 La fonction politique de la mesure

La mesure est donc bien inscrite dans des contextes et enjeux sociaux. L'utilité de la mesure doit donc se comprendre relativement aux enjeux des acteurs. Juhel & Gilles (2011) considèrent avec raison que *« Le véritable problème, scientifique et éthique, est non pas dans l'instrument lui-même mais bien dans la manière dont il est utilisé »* (p. 25) ; même si je pense qu'il y a aussi le problème de l'instrument comme j'ai essayé de l'expliquer : la manière dont est construit le test. Mais ce problème d'instrument ne peut être séparé du problème du contexte : la manière dont est utilisé le test.

Andronikonf affirme : « *Moi, je crois que ce sont toujours des échelles ordinales. Ce sont des jugements comparatifs de valeur qui se répercutent au niveau sociologique général, dans une société qui a besoin de hiérarchiser, à l'école ou dans l'entreprise, au niveau de la microsociologie et au niveau des jugements de valeur que l'individu porte sur lui-même.* » (Juhel & Gilles, 2011, p. 48). De même Cromby & Willis (2014) ou Desrosières (2014) pointent le lien entre la politique néolibérale est la généralisation des outils psychométriques par la politique gouvernementale ou au sein des organisations ; par exemple les échelles de personnalité usuellement utilisées en entreprise alors que les fondements scientifiques sont plus que critiquables (cf. tout le travail de Lamiell par exemple). En France, cette quantification généralisée s'opère par exemple par « les livrets de compétence » scolaires, ou par la mesure permanente de chaque individu via des évaluations professionnelles. Nous assistons à une production de mesures dont la base objective est parfois/souvent critiquable et qui ne peut se comprendre que par sa fonction dans le discours. Lacot et al. (2015) disent, avec raison : « *Consequently, the measurement issue in psychology, and particularly in psychological testing, is primarily ideological.* » :

En effet, le score s'inscrit dans une calibration qui se base sur un comparatif de personnes. Cela ne serait en soi pas un problème si la société considérait comme « normal » d'être plus ou moins habile dans différents domaines, si le cadre social ne créait pas cette concurrence de chacun contre tous. La fonction sociale des mesures ne découle pas du protocole de la mesure mais de la manière dont le cadre social actuel l'utilise comme fonction hiérarchisante permanente : notes scolaires, salaires, diplômes, évaluations professionnelles, ... Une société inscrite dans l'individualisme, dans une production inégalitaire permanente, ne peut justifier l'objectivité de cette hiérarchie que par des mesures (« évaluations ») considérées « objectives ». La science servira de recours à cette hiérarchie pour lui donner cette forme « non critiquable ». Ce n'est donc pas le score en lui-même qui est porteur de ce sens, mais la manière dont il s'inscrit dans un rapport social : l'évaluation concurrentielle permanente. Comme le disent Bertrand, Ahmadi, & Heuchenne (2008) : « *Le réalisme invoqué est en fait un appel à la soumission à la demande sociale d'un savoir fondateur de l'évaluation* » .

Je me retrouve donc dans cette critique de Vautier (2015f), qui différencie « mesure » et « évaluation ». Une mesure n'est pour lui que le résultat d'une « science objective », « l'évaluation » est dans le sens « interaction sociale » : « *Le terme d'évaluation possède une ambiguïté descriptive et appréciative redoutable. On dit qu'on évalue l'intelligence ou la taille d'un enfant. La taille est un attribut mesurable. D'où la tentation de conclure que l'intelligence est mesurable, puisqu'on l'évalue. Or les tests d'aptitude ne sont pas faits pour*

déterminer une quantité d'intelligence, mais pour assigner une place à l'enfant (via sa performance) dans une échelle de scores. En soi, un score n'est pas un jugement de valeur, mais ce n'est pas non plus le résultat d'un mesurage (Vautier, 2014d). Le score, interprété comme une propriété de l'enfant (le score traduit la performance, laquelle est le produit de l'intelligence, donc le score décrit l'intelligence), constitue la condition de possibilité pour que la formulation d'un jugement de valeur sur l'enfant, en fonction du contexte dans lequel il s'agit de l'insérer, acquière une factualité suffisante [...] "Insérer une personne dans un contexte social", en l'occurrence l'enfant évalué, implique une construction de significations à propos de la personne qui devient objet d'attention, objet à spécifier, à positionner et vis-à-vis duquel se positionner, objet à insérer dans un réseau d'enjeux relationnels et/ou institutionnels, le plus souvent implicites. ».

2.5.3.3 La fonction de légitimité de l'action

La construction et l'usage du test sont donc fondamentalement inscrits dans des enjeux sociaux : que mesurer ? Pourquoi ? L'usage social des mesures renvoie ainsi à des questions sociales. Desrosières a beaucoup travaillé sur le poids symbolique des valeurs numériques, et donc sur l'utilité de telles valeurs comme argument discursif. L'action proposée par un psychologue apparaîtra comme basée sur des faits objectifs s'il peut exhiber des mesures plutôt qu'un discours littéraire.

La mesure a donc une fonction sociale, donner une légitimité à l'action qui sera entreprise, comme le dit Bourdieu (2001, pp. 97–98): « *L'usage de formulations mathématiques abstraites affaiblit la tendance à percevoir la matière en terme substantiels et conduit à mettre l'accent sur les aspects relationnels.* ». En effet, en construisant des modèles statistiques avec des variables, on construit une vision relationnelle : « *la notion de variable a, dans la sémantique mathématique, une connotation différente de celle de nombre. Elle implique la formalisation de relations* » (Desrosières, 2008a). C'est la forme « relationnelle » (qui agit sur quoi ?) plus que substantielle (qu'est-ce que je mesure ?) qui domine la mesure en psychologie. Cette domination du relationnel répond à la fonction politique, à l'action à envisager, que porte la mesure... avec en conséquence un transfert du « *style de raisonnement des sciences de la nature vers les sciences sociales* » (Desrosières, 2008a, p. 69).

Or, comme je viens de le rappeler, la pratique concrète de la modélisation statistique n'offre pas plus de neutralité à l'interprétation que les études qualitatives. A ne pas discuter de la substance de la mesure (que mesure-t-on ? Pourquoi ?), ni des modèles statistiques (choix et

heuristiques afférents), mais à simplement baser la justification sur des corrélations, on construit un trompe-l'œil, un *fétiche* au sens donné par Vautier, dans la démarche d'objectivation de la réalité par une mesure (« *Un fétiche est un objet qui permet d'évoquer, ou d'invoquer si on s'adresse à elle, une entité métaphysique par le seul biais d'une convention ; il suffit que (presque) tout le monde croie à l'entité et croie que l'objet est relié à l'entité pour que cet objet devienne un fétiche* », Vautier, 2015c). Ce fétiche a une fonction : pouvoir effectuer les actions préconisées sur la base d'une objectivité scientifique... largement critiquable.

Mais, à l'opposé, refuser de quantifier n'est pas forcément neutre. Lorsque le ministre de la santé dit « *La souffrance psychique n'est ni évaluable, ni mesurable* » (Douste-Blazy : discours du 4 février 2005, Paris), ce discours n'est pas sans visée politique (le ministre s'adressait à une assemblée de psychiatre). Il faut donc décrypter dans chaque mesure sa pratique concrète, c'est-à-dire sa fonction dans l'espace social.

La fonction des « mesures » et « quantifications » dans l'espace public, les modifications de langage que cela entraîne, la prétention à l'objectivité au-dessus de la subjectivité, leur rôle politique, sont analysés finement par Desrosières, notamment dans Desrosières (2008a). Ce livre traite de la quantification socio-économique, mais « l'argument statistique », comme argument de « vérité » s'appuyant sur une épistémologie proche des sciences de la nature pour des champs des sciences sociales, peut être appliqué à la psychologie en tant que champ d'intervention sociale, avec ses propres outils de « mesures » statistiques ayant les mêmes fonctions « politique » (action) et basé sur la même épistémologie issue des sciences de la nature (« la vérité des modèles »).

2.5.3.4 Critique de la « boîte noire »

Mon discours n'est donc pas, je le redis, de s'interdire de mesurer, mais bien d'inscrire la mesure dans une démarche pragmatique, c'est-à-dire de discuter du contexte, enjeu, intérêt, validité, interprétation, d'une mesure (ou d'un modèle). Discuter de cela revient à considérer que bien des mesures usuelles sont contestables. Contestables dans leur fondement théorique, contestables dans leurs fondements statistiques, contestables dans leurs usages. Il faut donc pour un psychologue appréhender la mesure comme tout discours : de façon critique, c'est-à-dire questionner tout le processus de construction et d'usage de ce discours.

Or, c'est la grande difficulté de la mesure car il est difficile de remonter à la source du modèle, ses fondements idéologique et théorique, sa robustesse statistique, les choix

arbitraires de modèles effectués, ... La mesure protège celui qui va agir car elle est bien plus difficilement critiquable qu'un discours littéraire. Il faudrait construire, dans l'usage de l'espace public des mesures, un cadre critique de ces mesures pour éviter leur dimension « policière » : « *il faut défaire les boîtes noires pour montrer ce qu'elles dissimulent. La production statistique résulte de rapport de force. Elle est idéologique, policière [...] La statistique est bien d'abord, à travers ses objets, gage conventionnel de référence, dont l'existence permet à un certain type d'espace public de se déployer, mais dont le vocabulaire et la syntaxe peuvent eux-mêmes être débattus* » (Desrosières, 2010, p. 413). Comme le dit Gori (2014), il faut « *Réinsérer le chiffre dans un récit, dans un espace d'échange et de contestation [...] Le chiffre doit être une réalité transactionnelle, qui est produit et construit et dont on peut parler [...] Trouver dans la notion de valeur la capacité de parler le langage de l'humanité [...] Les indicateurs quantitatifs conduisent à une société de résignation.* ».

Or, ce discours critique de la mesure est peu présent car la mesure répond à une demande sociale, donc qui doit être exempte de critique, créant ce que Vautier appelle une ingénierie de fabrication d'avatars, au sens d'un dual artificiel de la personne, ayant pour fonction de positionner les personnes relativement à des objectifs : « *The second aspiration type is political: psychology hopes to become an academic discipline that is capable of not only satisfying but also creating a social demand, implying that it respects the way social problems, particularly psychological assessment problems, are proposed by the employers of psychologists* » et « *an engineering of avatar fabrication, i.e. representations of people in a virtual world.* » (Vautier et al., 2012). Or, je suis fondamentalement d'accord avec Vautier lorsqu'il écrit « *Il serait donc faux de considérer que la démarche habituelle est la seule possible pour un chercheur qui prétend contribuer à l'avancement des connaissances en psychologie générale* » (Vautier, 2015g).

2.5.4 Evaluer et comprendre

La fonction socio-politique de la mesure renvoie en miroir sa fonction d'aide à la compréhension. Je développe la critique de Vautier, marquant mes accords et désaccords avec lui afin de clarifier ma position critique.

2.5.4.1 Différence entre évaluer et comprendre

Vautier (2015f) pose une critique radicale de la mesure comme opposée à un processus de « compréhension » des dynamiques sous-jacentes : « *La dynamique de l'examen*

psychologique des aptitudes repose sur deux aspirations téléologiquement distinctes, évaluer vs comprendre, potentiellement incompatibles [...] On ne peut alors pas évaluer et comprendre la performance dans le même mouvement. Pour l'évaluer, il faut la scorer, c'est-à-dire la faire littéralement disparaître sous le nombre, lequel tirera sa signification d'une référentialisation que Danziger (1990, 1987) qualifie de galtonienne – l'étalonnage du score, ou encore le rapport à une distribution de référence. Tandis que pour comprendre la performance, il faut, adoptant une posture expérimentale, en découvrir les tenants – variables indépendantes – ce qui suppose de s'appuyer pleinement sur le langage descriptif qui permet d'identifier les changements intervenant au niveau de la variable dépendante. Du point de vue temporel, le calcul du score et son interprétation normative (ou, de manière synonyme, évaluative) prennent un instant, tandis que l'investigation de ce qui et de ce que détermine la performance est un effort de pensée, qui conduit peut-être à quelques spéculations ou hypothèses dans le cadre même de l'examen, ou d'avantage si on dispose de connaissances générales pertinentes – sinon, pas plus, parce que les personnes qui viennent se faire évaluer n'entrent pas de ce fait dans un programme de recherche particulier. ».

2.5.4.2 Evaluer pour comprendre

Ce que dit Vautier rejoint la forme « relationnelle » induite par des variables vs la forme « substantielle » (Bourdieu, 2001), notamment quand la « mesure » devient « évaluation » (donc s'inscrit dans la concurrence et normalisation inter-individus). Cependant, je ne suis pas entièrement en accord sur ce passage de Vautier car « scorer » peut servir à comprendre si on comprend cela dans une dynamique autre que celle actuelle. Ce que Vautier pointe via Danziger, ce n'est pas le processus d'objectiver une réalité par un symbolisme (une mesure par exemple), mais l'usage social de ce symbole. Je ne vois pas en quoi un score ne pourrait pas être « descriptif » et inscrit dans une démarche de donner du sens. L'écriture littéraire effectue le même processus : décrire pour comprendre. Dans un autre cadre social qui n'utiliserait pas de scores pour hiérarchiser ou pour les rapporter à une norme considérée comme impérative, cette représentation numérique pourrait servir dans une démarche « descriptive ». En effet, en quoi une description non numérique est-elle fondamentalement inscrite dans une démarche temporelle différente ? Toute description est « un instant » (« *Paul a du mal à réaliser certaines opérations arithmétique* » par exemple pourrait se transcrire par « *Paul a eu un score de 4 sur un test d'habileté arithmétique* »). C'est le sens de cette description, c'est-à-dire la démarche de remonter du symbole (verbal ou numérique) à la démarche dynamique

d'analyse. La différence est qu'il est intellectuellement difficile de réifier une phrase comme un fait « objectif » ; par contre les scores sont réifiés comme des objets propres à la personne. C'est bien l'usage social qui est critiquable.

C'est le poids symbolique des résultats statistiques, liés à la boîte noire associée, la dynamique sociale et politique cherchant à attribuer des faits objectifs aux individus pour justifier cette dynamique (licenciement, augmentation de salaire, orientation des certains domaines scolaires, ...) qui expliquent à mes yeux la tension entre « évaluer » et « comprendre » car les scores sont socialement réifiés comme des « vérités » pour agir et non pour comprendre, mais cette fonction sociale du score n'est pas intrinsèquement liée à la démarche de mesurer.

2.5.4.3 Critique de la pratique

Vautier critique le travail de Juhel & Gilles (2011) de la sorte : « *Le fait que la communauté des praticiens de l'évaluation psychotechnique cherche une légitimation de cette pratique dans la doctrine de la validation des tests (e.g., Juhel et al., 2011) constitue ainsi un obstacle épistémologique (Bachelard, 1983) au progrès de la connaissance scientifique de ce qui détermine la réponse aux items de tests – on pourrait utiliser la notion de dissonance cognitive.* ». En réalité, je devine une tension dans le texte incriminé par Vautier (Juhel & Gilles, 2011). Les auteurs développent toute une réflexion critique qui relativise les tests... pour par ailleurs leur donner une forme de légitimité « pratique ». Mais ce n'est pas la démarche de mesurer qui est un frein à la science (et encore faut-il s'entendre sur le terme « science »), c'est l'usage de la mesure.

Ce que l'on peut critiquer, c'est la systématisation de la mesure pour comprendre, comme seul mode opératoire : « *Tout se passe comme si, pour éviter la dissonance, la doctrine de la validation des tests s'était organisée pour ne pas accuser réception de la non mesurabilité des grandeurs psychologiques par des réponses à des tests ; la mesurabilité des grandeurs psychologiques tient le rôle de postulat fondateur, et la signification des notions de mesurage, de validité et de généralité grandeur* » (Vautier, 2015f)

Comme je l'ai déjà dit, c'est la certitude d'avoir construit un test valide indépendamment des enjeux et du contexte qui à mon sens pose problème car ce test rentre en contradiction avec notre compréhension en évolution des compétences (sociales). De plus, bien des tests sont basés sur une épistémologie « constructiviste » et ne peuvent prétendre refléter « une réalité » (c'est le test qui crée le concept, on quantifie par un index mais on ne mesure pas une réalité).

En ce sens, je suis d'accord avec Vautier pour dire que les tests actuels sont souvent de la psychotechnique « *c'est-à-dire un art de préparer des jugements évaluatifs selon les contextes dans lesquels on fait appel à leurs compétences* » (Vautier, 2015f).

Vautier (2015f) dénonce en conséquence la mythologie des mesures : « *une telle lucidité dans la communauté des psychologues de tous bords libèrerait la psychologie scientifique de l'impératif de fonder sinon tous, du moins une grande partie de ses programmes de recherche sur la mythologie des grandeurs psychologiques. Si la psychométrie a besoin d'une telle mythologie pour tirer profit de ses modèles à variables latentes, la recherche scientifique en psychologie n'a pas nécessairement besoin des modèles psychométriques. Elle a essentiellement besoin du droit à rendre compte de l'immensité de notre ignorance de ce qui cause les comportements qu'on sait observer* ». Je crois que la mythologie des grandeurs est la conséquence d'un autre mythe : « il existe des propriétés mentales fixes ». Mesurer, c'est se donner un cadre de représentation efficient, autant que l'objet mesuré puisse être considéré comme réel, et que la mesure puisse être considérée comme une représentation de l'objet. C'est donc l'objectif social de la mesure qui doit être questionné, mais en amont sa réalité non moins sociale.

De même, « notre ignorance » n'est une ignorance que si on croit qu'il y a une vérité cachée. Nous ne mesurons pas des propriétés mentales intrinsèques à une personne indépendante du contexte et de l'interaction sociale. Dans ce sens, la réalité est perceptible aujourd'hui et on se la représente éventuellement par une mesure. On peut critiquer, comme Ziegler & Vautier (2014), la naturalisation des scores (« *attempt to naturalize the values we want to assign to persons* »). Mais cela dépend du sens de "naturaliser". Si cela renvoie à une démarche comme en science « dure », à vouloir mesurer de « vrais objets » fixes/biologiques, en effet la naturalisation est critiquable... et c'est bien cette démarche qui domine. Mais si on entend « naturaliser » au sens de Dewey (Dewey, 1925/2012), c'est-à-dire comprendre notre objectivation comme un processus relatif à des objets réels, évolutifs, et donc objectivés relativement à un contexte, la « naturalisation » (au sens d'une réalité) ne me paraît pas problématique... pour autant qu'elle soit clairement inscrite dans un contexte. Cometti, dans son introduction à l'ouvrage de Dewey, pointe le fait que le terme « naturaliser » a un sens différent pour Dewey et l'empirisme-réaliste. Donc si on entend « naturaliser » dans le sens commun, je suis en accord avec Vautier, car il y a nettement un glissement des sciences sociales vers les sciences de la nature, comme si ces dernières devaient être le référent de notre démarche scientifique (Desrosières, 2008b), alors qu'il faut poser une épistémologie et méthodologie particulière pour les objets auxquels s'intéresse la psychologie.

2.5.5 L'exemple de l'intelligence

Je vais discuter de l'exemple de l'intelligence, je ne m'étendrai pas sur cette discussion car j'ai déjà fait référence à cette mesure précédemment. De plus, cela demanderait des développements plus longs. L'objectif n'est pas de polémiquer avec les auteurs qui développent des travaux sur la mesure de l'intelligence, mais de poser ma considération sur ce sujet relativement à quelques écrits pour inscrire la mesure de l'intelligence dans la critique de ce que je viens de développer.

En effet, beaucoup de travaux discutent de la mesure de l'intelligence, allant pour certains sur un retour vers la phrénologie (Bartholomew et al., 2009). Je discuterai sur la base de travaux récents publiés dans le livre collectif « *différences et variabilités en psychologie* » (Juhel & Rouxel, 2015), qui ne représentent pas un spectre exhaustif de toute la littérature, mais qui ont l'avantage d'avoir été publiés en 2015 en langue francophone, (dont avec un cadre culturel proche du mien, car, comme le rappelle Desrosières (2008b), l'intelligence est culturellement appréhendé différemment en France et aux USA par exemple) et avec différents chapitres discutant de la mesure de l'intelligence.

2.5.5.1 La mal-mesure de l'intelligence

Lecerf & Kieng (2015) font un état de l'art sur la mesure de l'intelligence. L'intelligence est définie de façon « consensuelle » pour ces auteurs comme « *capacité d'adaptation à des situations nouvelles, capacité qui permet de comprendre, d'apprendre* » même si Juhel (2005) rappelle que « *Les tentatives effectuées pour aboutir à une caractérisation plus précise de ce qu'est l'intelligence n'ont jamais rencontré beaucoup de succès* » (p. 3).

Cette définition donnée par Lecerf et Kieng semble de prime abord protéiforme puisqu'elle renvoie à des concepts différents : adaptation – compréhension – apprentissage. Or, comme je l'ai développé, si nous catégorisons par des concepts différents des objectivités, nous devons analyser différemment ces catégories car elles ne représentent pas une objectivité unidimensionnelle dans notre appréhension de la réalité. De plus, l'adaptation elle-même peut être vue de façon protéiforme. Des auteurs ont par exemple défini une « intelligence émotionnelle » (Mayer & Geher, 1996) relative à l'adaptation à ses émotions. Des auteurs développent des formes d'intelligence très variées discutant elles aussi de « l'adaptation » (relationnelles, ...) comme Gardner (2008). La littérature abonde ainsi de définition variées et contradictoires. Rozencwajg (2015) discute de l'approche intégrative de l'intelligence rappelant que des facteurs motivationnels ou émotionnels influencent les stratégies des

personnes (donc leur adaptation-compréhension-apprentissage). C'est le spectre des théories sur l'intelligence qui aujourd'hui nous oblige à l'appréhender d'une façon multidimensionnelle.

Ainsi, sur la mesure de l'intelligence, Lecerf & Kieng (2015) montrent dans leur revue de littérature que l'intelligence peut être appréhendée par un facteur unique, ou quatre facteurs, ou cinq facteurs, ou six facteurs et jusqu'à 16 facteurs, et en ne prenant que deux modèles théoriques de mesure de l'intelligence (WICS-IV et CHC). De plus, comme ces auteurs le disent (parlant de Wechsler) « *Son interprétation [...] a beaucoup évolué depuis le WB-I* » et le test actuel WISC-IV « *est destiné à refléter les conceptions contemporaines* ». Cette évolution du cadre théorique rentre en écho avec notre affirmation de la nécessité de renégociation... à la différence que Wechsler semble renégocier « *les conceptions contemporaines* » pour construire un outil censé mesurer toujours le même objet, « l'intelligence », comme propriété mentale trans-historique et trans-sociale ; comme si ses différentes moutures de test mesuraient toujours la même propriété mentale, donc indépendamment justement de la renégociation de ces propriétés telles qu'appréhendées socialement aujourd'hui vs hier.

Mais plus fondamentalement, comme je l'ai déjà dit, les tests d'intelligence ont été construits sans bases théoriques. O. Martin revient dans son ouvrage sur la genèse des tests d'intelligence au début du X^{ème} siècle, montrant la séparation nette entre le développement des outils quantitatifs et le développement de la théorie psychologique sur l'intelligence, les tests se sont développés non seulement sans recours au cadre théorique mais en contradiction avec celui-ci. « *Il semble exister une séparation franche entre les pratiques expérimentales des tests d'intelligence et les réflexions sur la nature de l'intelligence [...] D'un côté il existe des pratiques qui reposent implicitement sur des propriétés d'unidimensionnalité et de quantité de l'intelligence [...]; d'un autre côté, l'intelligence est affirmée comme étant multiforme, complexe et en tout cas non réductible à une seule grandeur quantitative* » (O. Martin, 1997, pp. 61–63).

Lecerf & Kieng (2015) rappellent ainsi « *l'absence d'ancrage théorique* » du test de Wechsler. Le concept de l'intelligence par Lecerf & Kieng (2015) ne renvoie donc absolument pas à une démarche réaliste. Nous ne sommes pas partis de « manifestations » qui nous semblaient pouvoir venir d'un domaine unique, nous n'avons pas objectivé ces manifestations par des « catégories homogènes », nous n'avons pas questionné si ces catégories pouvaient être représentées sur un continuum numérique. Il y a eu fondamentalement une démarche constructiviste, et on ne peut pas prétendre « mesurer » mais

juste « quantifier » (Desrosières, 2008b), sans pouvoir prétendre considérer comme réelle cette quantification.

2.5.5.2 L'orientation idéologique de l'intelligence

Si Wechsler a construit un test indépendamment d'un cadre théorique, il ne l'a pas fait en dehors d'un cadre idéologique. Gould (1997) revient sur la construction des tests d'intelligence, montrant leur ancrage dans une idéologie précise et avec des objectifs précis : racistes (les noirs sont moins intelligents que les blancs), sexiste (les femmes sont moins intelligentes que les hommes²⁸), eugénistes (il faut épurer la population des moins intelligents). Toute la mesure de l'intelligence a été initiée clairement dans un cadre idéologique : « *les biométriciens eugénistes anglais de la fin du XIXème siècle [Galton et Pearson entre autres] [...] sont d'abord des militants d'une cause politique, héréditariste et méritocratique* » (Desrosières, 2010, p. 402). Le livre « *the Bell curve* » (Herrnstein & Murray, 1994), qui re-popularisa l'idée d'une intelligence innée, est inscrit dans cette même lutte idéologique clairement définie « *héréditariste et méritocratique* » (je renvoie aux articles de Gould (1997) à ce sujet à la fin de la dernière édition de son ouvrage « *La mal mesure de l'homme* »). La dynamique de la mesure de l'intelligence est clairement dépendante de l'évolution de nos théories, de nos pratiques, mais aussi de nos idéologies au sens politique. Le débat entre Spearman et Thurstone sur le facteur unique vs une intelligence multi-facteurs est clairement inscrite dans la volonté de pouvoir hiérarchiser les individus à partir d'un score unique vs l'idée d'une vision plus complexe et protéiforme des individus.

Il y a, dans la démarche de soit-disant mesure de l'intelligence, en réalité une construction d'un indicateur quantitatif n'ayant qu'une base idéologique (démarche constructiviste), et absolument pas une « mesure » d'une ou de propriétés mentales (démarche réaliste). Comme le disent Lecerf & Kieng (2015) « *un test d'intelligence est toujours l'opérationnalisation d'une conception théorique* »... et je rajouterai « idéologique », car la théorie sur l'homme n'est jamais séparée d'une vision politique de l'organisation sociale. Mais « *l'opérationnalisation* » d'un indicateur ne le transforme absolument en « mesure » au sens d'une épistémologie réaliste.

²⁸ Gould (1997) et Hacking (2008) rappellent par exemple que les premiers tests de Q.I. ont été épurés des items où les femmes avaient de meilleurs scores que les hommes car les femmes « *sont par nature moins intelligentes que les hommes* ».

2.5.5.3 L'invariance idéologique de l'intelligence

L'autre débat, tel que mené par les auteurs de « *The Bell curve* » ou Pearson et bien d'autres, est sur l'inné vs l'acquis, qui est aussi fondamentalement un débat à résonance politique (faut-il aider les moins intelligents si leur intelligence n'est pas modifiable ? Car la question des objectifs des dépenses gouvernementaux étaient clairement l'objectif politique de « *The Bell curve* », voir les commentaires de Gould). Cette discussion renvoie à l'invariance de l'intelligence chez les individus.

Lecerf & Kieng (2015) discutent d'une invariance de l'intelligence chez les personnes : « *le niveau intellectuel d'un individu devrait rester stable au cours du temps* » s'appuyant sur les travaux de Deary, Whalley, Lemmon, Crawford, & Starr (2000) et sur leur propres travaux. Notons déjà que la formulation « niveau intellectuel » ne renvoie pas de façon évidente à une intelligence adaptative comme ces auteurs l'ont définie initialement. Le niveau intellectuel est clairement une notion inscrite socialement dans une hiérarchie de compétences telle que notre société la promeut. Les résultats de Lecerf & Kieng (2015) sont fragiles (96 enfant avec deux répliques séparées de 1 an ½ à 3 ans) et montrent que cette stabilité n'est pas uniforme pour toutes les dimensions ; et au niveau intra-individu, un tiers des enfants ne montrent pas de stabilités (résultats de Lecerf & Kieng, 2015). Le travail de Deary et al. (2000) renvoie une corrélation de 0.63 sur 101 personnes analysées à 11 ans (1932) et à 77 ans (1998), donc avec une perte de 90% de la cohorte de 1932. Cette corrélation peut être interprétée de différentes manières, on peut dire que pris isolément le score de 1932 « n'explique » (au sens du R²) que 40% du score de 1998, donc que 60% des variations inter-individuelles sont dues à des facteurs exogènes. De plus, une telle stabilité relative pose la question de la fonction du score de 1932 dans la poursuite de la vie sociale des jeunes scorés en 1932, de la qualité du test de 1932 répliqué 66 ans après, ...

Bref, il y a ici, comme dans toute analyse statistique, des choix d'interprétation des résultats. Ces choix ne sont pas neutres lorsque l'on parle de « l'intelligence » car ce concept est structurant pour hiérarchiser les personnes sur un continuum de façon « scientifique ». Je soupçonne Deary de ne pas être en dehors de toute démarche idéologique car un autre article qu'il a co-signé renvoie clairement à la phrénologie (Bartholomew et al., 2009). On retrouve dans les deux articles de Deary le cœur de l'idéologie à l'émergence de la mesure de l'intelligence : l'intelligence est innée (donc non modifiable) et liée à la taille du cerveau. A l'inverse, De Ribaupierre (2015) rappelle les écarts intra-individuels au test de WISC-IV qui

contredisent l'invariance d'une personne : « *l'hétérogénéité, c'est-à-dire une variation de notes d'un sous-test à l'autre, est la règle plutôt que l'exception* » (p. 162).

Gould (1997) rappelle la genèse des tests d'intelligence où nous sommes passés de son initiateur (Binet, qui voulait apporter un indicateur pour aider les enfants dans leur scolarité) à une mesure hiérarchisant les personnes (Spearman et tous les développements qui ont suivi) clairement inscrite dans une idéologie cherchant à justifier les inégalités sociales par les inégalités intellectuelles (voir aussi l'ouvrage de O. Martin, 1997). Donc « l'invariance » de la mesure d'un test est une hypothèse forte portée par Spearman et d'autres... que l'on retrouve actualisée par Deary et al. (2000). Gould propose d'autres interprétations des données « homogènes » (autant qu'elles le soient), posant explicitement ses analyses comme reliées à son positionnement idéologique : l'influence de facteurs sociaux. « *La science est enracinée dans l'interprétation créatrice [...] ceux qui interprètent [les chiffres] sont souvent prisonniers de leur propre raisonnement. Ils croient en leur propre objectivité et ne parviennent pas à percevoir le préjugé qui les incite à choisir une interprétation parmi toutes celles qui pourraient également être compatibles avec les chiffres qu'ils ont relevés* » (Gould, 1997, p. 108).

2.5.5.4 Critique de la mesure de l'intelligence

Les résultats empiriques de Lecerf & Kieng (2015) montrent des convergences et divergences de données recueillies relativement aux différents modèles. Mais tous ces résultats 1/ sont basés sur des corrélations, en quelque sorte via des analyses factorielles qui souffrent du problème de la robustesse de telles analyses comme déjà discuté ; 2/ sur des résultats moyens, donc non applicables aux singularités. Or, il y a une tendance dans tous ces travaux à considérer que des corrélations « valident » la mesure de l'intelligence, avec clairement un biais issu des choix (items retenus, nombre de facteurs retenus, ...). Cela n'invalide pas les résultats empiriques, mais oblige à un retour réflexif sur ces résultats comme toute analyse quantitative.

Or, toutes ces analyses révèlent à mon sens un fantasme récurrent qui est la quête de l'intelligence en tant qu'entité « véritable », volonté de pouvoir mesurer un objet réel dans le cerveau identique en tout temps et tout lieu, comme si l'intelligence était une propriété biologique et non une propriété mentale dans un cadre social précis. Car il y a bien un ancrage biologique dans toute la littérature passée et présente sur l'intelligence. Lecerf & Kieng (2015) rappellent par exemple que le test CHC est construit relativement à la recherche

neuropsychologique. Il y a clairement une démarche inverse de ce que devrait faire la psychologie : partir de faits/manifestations pour construire des outils de compréhension (quantitatifs ou non) de propriétés mentales comprises comme des propriétés inscrites dans une interaction sociale. Pour l'intelligence on part d'un concept affirmé, et on essaye ensuite de lui donner une forme de réalité par une mesure (qui est en fait une quantification) avec des sous-bassements idéologiques et épistémologiques.

On peut donc apporter à la mesure de l'intelligence deux critiques fondamentales portées par bien des chercheurs : 1/ une critique métrologique : les mesures d'intelligence deviennent des « tautologie » (Juhel & Gilles, 2011) car l'intelligence est ce que le test mesure (car les fondements théoriques sont faibles) ; en conséquence, les résultats empiriques montrent que des modèles différents validant des données similaires et les résultats empiriques sont bien souvent contradictoires (voir tous les résultats apportés par Lecerf & Kieng, 2015), contredisant une propriété mentale unidimensionnelle ou multi-dimensionnelle clairement définie ; enfin les modèles utilisés sont posés comme « hiérarchiques » entre les différentes formes d'intelligence, en contradiction avec une vision interactive de propriétés mentales comme je l'ai exposé lors de ma discussion sur les *Neural Networks*²⁹ ; 2/ philosophique : les tests d'intelligence sont fondamentalement une démarche constructiviste liées à la volonté de hiérarchiser des individus, avec un ancrage biologique ; voulant donner une caractéristique pérenne de la personne, sur une propriété considérée comme centrale pour justifier leur position sociale, indépendante des contextes sociaux et des enjeux de cette mesure.

2.5.5.5 Pour une mesure pragmatique

Comme la littérature le montre, on retrouve une hétérogénéité de caractérisation de l'intelligence. Je suis critique de la conceptualisation de l'intelligence comme une capacité unique intrinsèque... voire critique de ce concept « d'intelligence » en tant que mesure hiérarchique des individus (voir toute la discussion critique de Gould (1997) par exemple ou l'historique de ce concept chez Juhel (2005)). L'objectif de mesurer une propriété mentale intrinsèque à l'individu comme l'intelligence est un « fétiche » au sens donné par Vautier (2015c) qui a une fonction idéologique. Comme pour toute propriété mentale, il n'y a pas d'intelligence « en soi », il n'y a que des propriétés mentales contextualisées, répondant au

²⁹ voir par exemple le modèle simplement multiplicatif intégrant « l'intelligence », sans effet d'interaction et boucle entre propriétés mentales, de Bailleux, Dauvier, & Perret (2015).

cadre social de leur mesure. C'est bien le contexte social et les enjeux qui vont orienter notre appréhension de manifestations dites « d'intelligence ».

Mais si nous partons d'un concept d'intelligence opératoire, c'est-à-dire construit dans une perspective d'aider socialement les personnes, c'est-à-dire clairement inscrite dans un contexte et enjeu précis, nous pouvons considérer qu'évaluer des habilités dans une dynamique non hiérarchisante est potentiellement possible, par exemple habilité spatiale, habilité logique et habilité verbale (Markus & Borsboom, 2013a). Mais cette catégorisation est conséquente d'un cadre social car « *Les test sont socialement et culturellement conditionnés* » (Juhel, 2005, p. 14), qui explique l'effet Flynn (une croissance générale du QI dans les populations occidentales qui peut difficilement laisser considérer que les capacités intrinsèques des individus soient différents des époques antérieures). Dans une autre civilisation, l'intelligence pourrait être catégorisée en sous-dimension très différentes si on reprend le cadre de la définition « capacité d'adaptation ». Je ne sais plus quel philosophe a écrit qu'il ferait confiance à l'intelligence d'un polytechnicien pour gérer une centrale nucléaire ; par contre il préférerait s'en remettre à l'intelligence d'un indigène dans la forêt amazonienne car l'adaptation à ce milieu demande d'autres habilités que celles de notre société. Cela ne veut pas dire qu'habilité spatiale, habilité logique ou habilité verbale n'ont pas de réalité, mais une réalité « pour nous », dans notre contexte social. Il ne peut y avoir donc que renégociation de ces catégories et domaines uniques car il n'y a pas UNE intelligence ou DES catégories d'intelligence indépendamment de la pratique sociale. D'ailleurs, c'est ce qu'offre toute la littérature sur l'intelligence, c'est-à-dire cette renégociation permanente en fonction de l'évolution de nos catégorisations.

Vautier (2015f) critique par exemple le test WISC-IV : « *Par exemple, "l'indice de compréhension verbale du WISC-IV est une mesure de la formation de concepts verbaux, du raisonnement verbal et des connaissances acquises dans le propre environnement du sujet"* (Wechsler, 2005b, p. 89). *La formation de concepts verbaux etc. ne dénote pas une grandeur. Dans la phrase "l'ICV actuel peut être considéré comme une mesure affinée et plus pure du raisonnement verbal et de la conceptualisation que [...]"* (Wechsler, 2005b, p. 89), *le raisonnement verbal n'est pas une grandeur, pas plus que le raisonnement perceptif et fluide qu'on trouve dans la citation : "l'indice de raisonnement perceptif est une mesure du raisonnement perceptif et fluide, du traitement spatial et de l'intégration visuomotrice"* (Wechsler, 2005b, p. 89) ». La relation entre « formation de concepts verbaux » et « mesure » est en effet problématique ; non qu'une mesure soit en soi une représentation erronée d'une propriété, mais peut-on considérer pouvoir représenter sur un continuum numérique une telle

propriété ? Cela renvoie à ma discussion sur la protéiformisation car il y a plusieurs « idées » dans ce que dit Wechsler. L'habilité de « formation de concepts verbaux » est-elle assimilable à l'habilité de « raisonnement verbal » ou de « connaissance acquises dans le propre environnement du sujet » ? N'y-a-t-il pas ici des sous-dimensions qu'il faudrait affiner pour comprendre comme le demande Vautier ; et vérifier pour chacune d'elle si elle peut être considérée comme une propriété mentale représentable sur une échelle numérique ? Agglomérer par un score unique des « compétences » variées interdit effectivement de comprendre. Mais inversement, développer des sous-mesures de compétences variées, en analysant pour chacune d'elle sa représentation potentielle d'une façon ordinale ou numérique, ne me paraît pas contradictoire avec la recherche de sens.

Donc je me trouve en accord avec Vautier (q2015b) quand il dit : *« L'utilité de l'évaluation du retard ou de l'avance de développement peut être défendue sur une base sociétale et éthique (par exemple, il est bon que les enfants puissent être pris en charge dans des conditions éducatives adaptées à leurs performances intellectuelles — mais est-il bon que l'institution scolaire établisse, sous couvert de la science psychologique naissante, qu'un enfant est ou bien normal ou bien retardé ?). Pour autant, les psychologues ne devraient pas participer à la réification du niveau de développement intellectuel, qui, parce qu'elle repose sur la fausse idée de sa mesurabilité, conduit à une forme de scientisme sans lequel la naturalisation quantitative de la catégorie d'enfant retardé est impossible. »*

2.5.6 Critique de la science

La fonction sociale de la mesure en psychologie dans l'espace public renvoie aux objectifs idéologiques liés. Cette fonction s'appuie sur un discours scientifique qui donne la légitimité aux actions entreprises à partir des mesures. Il y a donc une nécessité d'un retour critique de la science dans sa fonction sociale. Ce retour critique de la science est mené par des courants de la sociologie de la science, car en reprenant Gould (1997) : *« Les déterministes se sont souvent servi du prestige de la science comme d'une connaissance objective, libre de toute influence sociale et politique [...] Je pense plutôt que l'on doit appréhender la science comme un phénomène social, comme une entreprise humaine dynamique »* (p. 52)

2.5.6.1 La boîte noire scientifique

Pour que la mesure puisse jouer une fonction socialement démocratique dans l'espace public, il faut que la transparence puisse exister, il faut que les mesures puissent être décortiquées

dans leurs fondements, ce que Desrosières appelle pénétrer « la boîte noire » (Desrosières, 2008b). Je me retrouve donc en parfait accord avec Vautier lorsqu'il écrit : « *La validité d'une proposition scientifique repose par principe sur la possibilité dont jouit quiconque de vérifier par lui-même la validité de cette proposition. Bien entendu, il n'est en fait pas possible à quiconque de vérifier tout le savoir scientifique. Mais il importe que le jugement de validité d'une proposition ne soit pas en droit réservé à une élite, et que la vérification soit bien l'affaire de chacun* » (Vautier, 2012). Vautier (2015c) exige donc, avec raison, la capacité du commun des mortel à pouvoir s'appropriier la « mesure » qui joue un rôle social : « *S'il est nécessaire, le scoring devrait être justifié dans le contexte d'une convention sociale non fétichiste, c'est-à-dire telle qu'il soit publiquement reconnu que le score n'évoque pas de propriété naturelle de la personne qui en est titulaire (son intelligence, son anxiété, etc.).* ».

L'ironie dans ce phénomène de « validité objective de la mesure » est que bien des scientifiques usant de modèles statistiques avec variables latentes n'ont guère de connaissance plus élaborées que le commun des mortels sur les modèles statistiques du fait du peu de culture mathématique qu'ils ont (Hacking, 2008). Bien de scientifiques en sciences sociales affirment par leurs travaux avoir obtenus des résultats scientifiquement valides en usant de modèles dont ils n'ont en réalité qu'une connaissance parcellaire (ou par simple mimétisme, reprenant ce que d'autres ont publiés comme modèle de référence) au sens de la statistique-mathématique.

Mais au-delà de la maîtrise technique, il y a une non-connaissance de l'histoire de la construction statistique, des enjeux et choix opérés qui s'inscrivent en phase avec l'histoire sociale (voir pour exemple les développements sur l'histoire d'outils statistiques rapportés par Armatte (2010) ; Desrosières (2008a, 2008b, 2010, 2014) ; Hacking (1992, 2002) ; O. Martin (1997) ; Michell (1999)). Comme le montrent tous ces auteurs, les modèles statistiques devenus populaires ne sont pas des modèles déconnectés des enjeux sociaux-politiques. Il y a nécessairement un retour critique à faire sur leur construction et usage. Hacking (1992) considère que la statistique, comme mode de raisonnement, a créé de nouveaux objets sociaux ; un nouveau modèle explicatif ; de nouveaux critères d'objectivité. Pour Hacking, la statistique est une procédure de légitimation « intersubjective », ni vraie ni fausse mais à considérer comme toute procédure de légitimation dans son cadre social et historique.

Il y a donc dans la science une double réalité, la réalité « interne » (celle de la communauté scientifique) et la réalité « externe » (la médiatisation/l'usage politique de la science). La science produit socialement un savoir qui de l'extérieur est inattaquable, sauf à se faire considérer comme inscrit dans une vision ascientifique/anti-progrès (Carnino, 2015). Mais de

l'intérieur, la science est beaucoup plus critique sur elle-même. Il y a cette forme d'hypocrisie dont parlait Bourdieu (2001, p. 152) : « *La vision officielle de la science est une hypocrisie collective propre à garantir le minimum de croyance commune qui est nécessaire au fonctionnement de l'ordre social [...] Tout le monde sait la vérité des pratiques scientifiques [...] et tout le monde continue à faire semblant de ne pas savoir et de croire que ça se passe autrement.* ».

2.5.6.2 La fonction politique de la science

Si la science fonctionne à deux niveaux, à un niveau interne de façon critique, et à un niveau externe avec l'image d'un savoir absolu, c'est que la science a une fonction sociale et politique. L'inaccessibilité au savoir scientifique par le commun des mortels permet d'user des résultats scientifiques dans leurs fonctions politiques. La science joue socialement et politiquement une fonction d'expertise pour l'espace public mais construite en dehors de l'espace public, comme un champ isolé apportant à la collectivité le « savoir » par essence inattaquable car « scientifique ». Cela oblige à un retour critique de « la science ».

L'histoire de la science statistique a été rapportée par les livres de Desrosières ou d'Armatte que j'ai déjà cités, inscrivant ce champ scientifique comme inscrit dans des connexions sociales et politiques. Mais au-delà de la statistique, l'émergence de « la science » en tant que champ unifié remonte au XIX^{ème} siècle, Carnino (2015) explique : « *en effet, comme une étude lexicale soigneuse le montre, l'idée contemporaine de "la science" n'existe pas réellement avant le XIX^{ème} siècle en Europe* » (p. 12). Carnino développe de façon convaincante l'émergence du terme « la Science » au XIX^{ème} siècle, voir le premier chapitre de ce livre et les graphiques d'occurrences des termes « science » et « scientifique » associés qui montrent l'émergence de ces termes au XIX^{ème} siècle ; et les chapitres suivants sur l'inscription de cette émergence dans le développement de la société avec ses icônes, comme Pasteur. L'émergence de « la science » est ainsi en lien social et économique avec le développement industriel, à l'inverse d'une vision d'une « science pure ». « *Le savoir scientifique fait l'objet d'une transaction : contre le service rendu, le savant bénéficie de la reconnaissance de tous. La science s'inscrit alors dans une rhétorique du don pour mieux en dissimuler les enjeux industriels et économiques.* » (Chauveau, 2014).

Il est souvent considéré que « *la science produirait le vrai grâce à sa méthode infaillible, fondée sur l'observation des faits - matériau éternel et irréfutable sur lequel s'arrime toute connaissance* » (Carnino, 2015, p. 9) ; or « *nombre d'opinions relatives à la science relèvent*

de la croyance [...] à la lueur des avancées récentes de l'histoire sociale et culturelle des sciences, il est aujourd'hui guère possible de continuer à penser la science comme un univers étanche et clos, qui flotterait au-dessus ou au-delà du social. Non seulement la science n'apparaît jamais concrètement comme une pure entreprise de connaissance qui serait parfois infléchie ou détournée par des bais extérieurs, mais même la science la plus "pure", celle qui trône au panthéon du savoir rationnel, alimente ses réflexions grâce à la religion, l'industrie, la politique, l'économie et la technique » (p. 10-11).

Le succès de « la science » ne s'explique pas seulement par sa capacité à répondre au développement économique. La « science » en tant qu'acteur social a une fonction idéologique fondamentale. Carnino apporte un regard critique sur cette émergence de « la science », avec une dynamique complexe comme il le dit lui-même, mais dont la ligne directrice peut se comprendre comme fonction politique en France à l'émergence de la 3^{ème} République dans la confrontation avec l'Eglise. Mais la « science » revêt, au-delà de sa fonction politique contre l'Eglise, une forme idéologique comme « *nouvelle religion* » ou « *nouvelle foi* » (avec ses saints/icones comme Pasteur, Galilée, Gutenberg, ... dont les statues en sont les symboles), permettant « *le report des griefs populaires à un avenir plus ou moins lointain [...] la science est le nouveau relais du progrès, elle fonde, le représente, le prouve et finalement s'y substitue aux yeux de tous les sceptiques qui refusaient jusqu'alors de se laisser convaincre. L'émergence de l'idée de science participe donc à l'acclimatation industrialiste des populations ouvrières, et constitue une forme de pacification idéologique* » (p. 206-210).

La fonction de la science devient un outil d'autorité pour le champ politique. Les acteurs politiques de la 3^{ème} République en avait une conscience aigüe (cf. toutes les citations données par Carnino : « *Cette perception du rôle structurant de la science infuse l'ensemble de la classe politique républicaine* » p. 205), ce qui autorise Carnino à considérer que « *Le dogme nouveau, le positivisme, pose la science comme doublement fondatrice du devenir des sociétés contemporaines : d'une part elle en constitue le socle spirituel, et d'autre part elle anime le mouvement naturel qui porte l'humanité vers le progrès par l'industrie. Ordre et progrès, c'est-à-dire science et industrie – telle est la religion du monde futur [...] la science constitue la preuve et la caution des transformations en cours* » (p. 240).

Cette fonction politique de « la science » a pour « *principale conséquence d'exclure la science du champ politique [...] nul ne peut plus intervenir directement dans les affaires de la science sans avoir été au préalable intronisé en son sein* » (p. 257). Cela explique que pour la psychologie, comme tout champ scientifique, « *l'enjeu épistémologique majeur de toutes les*

disciplines savantes consiste désormais à prouver qu'elles sont "scientifiques", sous peine d'être exclue du champ du savoir certain et indubitable » (p. 257) car si ce n'est pas une science, « *tout à chacun peut se permettre d'avoir une opinion sur le sujet* » ; or en science les « *jugements sont l'apanage des seuls élus qui la pratiquent-ils deviennent au propre comme au figuré parole d'évangile* » (p. 258). Ceci explique le désir constant de la psychologie à vouloir coller au plus près des sciences de la nature, notamment la physique, parangon de « la science ». O. Martin (1997) rapporte ainsi l'émergence de la psychométrie à la fin du XIXème siècle et le début du XXème siècle où « *Des démarches visant à établir la psychologie comme science rationnelle et quantitative vont s'affirmer* » (p. 2).

Ainsi la science devient un champ séparé et réservé aux seules personnes reconnues par elle, avec donc « *la parole d'expert* » qui « *est garantie de l'objectivité et de la neutralité de tout jugement* » et donc « *exclu de la sphère démocratique* ». La science devient une arme par son discours indiscutable, permettant des « *arguments d'autorité* », instituant « *un espace étanche à l'opinion du peuple* ». Nous retrouvons ces mêmes considérations sur l'autorité de la science dans l'espace public chez Armatte (2010); Bourdieu (1997); Desrosières (2014); Hacking (2008). Comme la science s'érige en experte du savoir, elle conteste le savoir extérieur à son champ. Putnam (1984) dénonçait ce positionnement philosophique scientiste : « *je suis en train de suggérer que l'apparition dans notre culture d'une tendance philosophique, obnubilée par la réussite de la science à un point tel qu'elle est incapable d'admettre l'existence du savoir et de la raison en dehors de ce que nous avons l'habitude d'appeler les sciences* » (p. 206). Nous retrouvons dans l'ouvrage « *histoire populaire des sciences* » de Conner (2011) cette histoire de l'accaparement du savoir par les seuls scientifiques, étouffant toute prétention au savoir « populaire ». Le seul savoir reconnu comme légitime est celui de la communauté scientifique extérieur à l'espace public : « *un effet d'autorité intervient. La science a découvert, de façon générale, comment les choses sont (c'est ce qu'elle nous dit), comment elles doivent être, dans l'état actuel des choses [...] [servant à] garantir l'autorité culturelle de la science* » (Hacking, 2008, pp. 132–133).

2.5.7 Conclusion

Cette fonction politique de la science permet d'appréhender les enjeux de la psychométrie car, comme le dit (Desrosières, 2014, p. 252) « *La mise en perspective de ces critiques de la quantification avec celle observées dans d'autres disciplines montre que seule l'analyse politique de leurs usages sociaux permet de dépasser un face-à-face purement cognitif* ». De

même, Vautier (2015b) critique la fonction de la psychométrie : « *Le scientisme psychotechnique tient à l'opacité des motivations subjectives de la convention qui édicte l'efficacité contre les faits falsifiant, opacité qui offre le support de projection suffisant pour que le profane face confiance aux experts qui se sont constitués en adoptant (et en défendant) cette convention.* ». Ainsi, les mesures dites « scientifiques » deviennent des enjeux dans l'espace public. La non transparence, l'incapacité de pénétrer « la boîte noire » de la modélisation statistique permet à la psychométrie de pouvoir être reconnue comme « science » et en conséquence de pouvoir user d'argument d'autorité (en résumé « *c'est prouver* »).

Or, les sous-bassements épistémologiques et méthodologiques de la psychométrie peuvent être critiqués... mais uniquement par des acteurs de son champ. La critique elle-même ne sort pas du champ de la science et ne pénètre pas plus l'espace public. La critique prend place presque uniquement dans ce jeu interne à la science, et ne critique pas la pratique « psychotechnique » dans l'espace public, se parant de l'autorité de « la science ». Armatte (2005, p. 103) rapporte en conséquence les critiques des modèles normatifs issus des modèles mathématiques, en citant Badiou « *La notion de modèle empirique en sciences sociales n'est pas autre chose pour lui qu'une poudre aux yeux ou un objet idéologique de manipulation.* ».

3 Conclusion de la partie 1

Je conclus cette partie épistémologique par une synthèse de ma thèse avant de revenir sur la littérature académique.

3.1 La thèse défendue

Je reprends ci-dessous les grandes lignes de la thèse défendue dans ce travail. Je reviendrai dans la conclusion finale sur des aspects plus généraux.

3.1.1 Ma thèse

La thèse que je défends considère que les propriétés mentales sont des réalités. Mais cette réalité doit être appréhendée dans les interactions sociales, c'est-à-dire au-delà d'une vision biologique de ces propriétés mentales. Non pas que ces propriétés mentales n'aient pas d'existence biologique, mais elles doivent se comprendre d'une part comme des propriétés liées à l'être social (et non l'être biologique) ; et d'autre part leur catégorisation est liée à la praxis, c'est-à-dire liée aux contextes et enjeux. En conséquence la conceptualisation des propriétés mentales ne peut se faire qu'au travers de manifestations dans l'interaction sociale. Je peux reprendre à mon compte les propos de Varela et al. (1999) « *ce livre diffère [...] en un point fondamental : il envisage la conscience non pas comme une simple capacité ou performance dont on cherche à rendre compte [...] mais avant tout comme une expérience* » (p. 10) et « *Nous affirmons avec Merleau-Ponty que la culture scientifique occidentale nous invite à considérer nos corps comme une structure à la fois physique et vécues – bref, comme à la fois « extérieures » et « intérieures », biologiques et phénoménologiques* » (p. 17). Ces auteurs renvoient, me semble-t-il, aussi à une démarche pragmatique proche de celle que je développe : « *quand c'est la cognition ou l'esprit qui est examiné, l'omission de l'expérience devient intenable, voire paradoxale* » (p. 40), et « *nous pouvons remarquer que ce point de vue est en résonance avec le pragmatisme [...] la relation du corps et de l'esprit est conçue en fonction de ce qu'elle peut accomplir* ». Donc considérant que l'on ne peut séparer le corps et l'esprit au sens du corps en pratique, comme je le considère aussi. Je me permets aussi de citer Besnier dans l'article du 12/10/2015 du Monde des Sciences (« *une machine peut-elle ressentir de l'émotion ?* ») : « *Ce sont les signaux qui comptent. En toute rigueur, votre*

intériorité, je ne sais pas si elle existe [...] Elle se manifeste à moi par des signes extérieurs. Je ne sais pas ce que c'est que de ressentir une émotion chez un autre être humain. Je peux ressentir de l'empathie, mais je peux aussi en ressentir face à un acteur, qui simule. Je ne peux pas savoir si ces signes émanent d'une intériorité ou d'une simulation ».

Comme les manifestations, la catégorisation est liée à une praxis, ce qui veut dire que les propriétés mentales ne sont à comprendre que comme des catégories à renégocier conceptuellement. Cela ne retire rien à leur réalité, mais « réalité pour nous ». Les manifestations révèlent une propriété mentale qui peut être représentée par une mesure. Représenter une propriété mentale sur une échelle numérique est possible, autant que l'on puisse justifier l'unidimensionnalité de la propriété, l'indépendance locale des manifestations, une densité empirique de la propriété. Cette représentation ne peut prétendre être la réalité, ce n'est qu'une réification, comme l'est toute conceptualisation et représentation. Le symbolisme mathématique épure en grande partie le discours de tensions culturelles, mais en psychologie le recours au langage mathématique ne sert qu'à formaliser des modèles statistiques. Ces derniers ne sont pas épurés de tensions culturelles/subjectives, et donc sont autant que le discours littéraire emprunts de choix et heuristiques. L'usage de modèles statistiques sert à donner du sens, mais en retour il ne peut prétendre apporter tout le sens. D'un résultat probable en moyenne on ne peut glisser vers une règle applicable à chaque singularité. La modélisation sert de repère normatif, mais ne peut prétendre révéler toutes les particularités.

En conséquence, autant je considère que la mesure d'une propriété mentale est pratiquement possible ; autant je suis critique de l'usage social des mesures (ou évaluations) en psychologie telle que pratiquée aujourd'hui. La mesure en psychologie renvoie aujourd'hui à une démarche inscrite dans un cadre épistémologie empirique-réaliste, servant de justificatif pour ériger des lois psychologiques comme des lois de la nature, avec une fonction normative et politique qui rentre en contradiction avec une compréhension critique de la démarche scientifique.

3.1.2 Le pragmatisme comme méthode

Juhel & Gilles (2011) considèrent (comme la plupart des auteurs) qu'il y a deux points de vue sur les propriétés mentales. Le premier dominant, celui que j'ai appelé empiriste-réaliste ; « *Le point de vue actuellement dominant considère que les modèles théoriques sont, au moins partiellement et approximativement, vrais, et qu'ils permettent d'atteindre des entités réelles, en attente d'être découvertes. Ce courant se réclame ainsi d'une certaine forme de réalisme*

scientifique » (p. 5). Ce point de vue semble interdire de « mesurer » des propriétés mentales car toute la critique de la mesure discutée précédemment laisse considérer qu'il n'y a aucun moyen de justifier une mesure quantitative d'une propriété mentale dans ce cadre épistémologique. Le second point de vue, que j'ai appelé constructiviste, considère que « *la réalité psychologique est inséparable du sujet et qu'elle ne peut être comprise du seul point de vue de l'observateur* » (Juhel & Gilles, 2011). Ce point de vue de façon symétrique interdit toute mesure car nous sommes clairement dans un processus instrumentaliste, donc on confond ici la mesure et le processus de mesurer.

Je positionne toute ma thèse relativement à une épistémologie pragmatique-réaliste, c'est-à-dire dans une épistémologie qui ne considère pas les propriétés mentales comme de simples constructions mais bien comme des réalités, mais par contre qui considère que la démarche scientifique doit être pratique, donc critique des négociations passées de propriétés mentales. Le cadre épistémologique pragmatique-réaliste proposé n'est pas un entre-deux entre le réalisme-empirique et le constructivisme, il propose une autre démarche scientifique. On peut retrouver cette nécessité d'un autre cadre épistémologique dans ce que Desrosières (2008b, p. 65) exprime du réalisme ou du constructivisme : « *quel langage imaginer qui ne soit ni celui du réalisme métrologique naïf des sciences de la nature (qui est comme le rêve perdu impossible des sciences de la nature), ni celui d'un constructivisme relativiste, vu comme la négation de la dure réalité d'un monde social dont la description ne relèverait que de l'arbitraire de rapports sociaux contingents orientés par des intérêts particuliers ?* ». Desrosières n'offre pas de réponse à sa question, et je crois que le pragmatisme-réaliste, tel que proposé, offre une réponse car il est une théorie de la connaissance qui évite les écueils de l'empirisme-réaliste et du constructivisme décrits par cet auteur.

Un tel cadre épistémologique rend possible la mesure de propriétés mentales, mais en critique des mesures usuellement effectuées qui sont le plus souvent en référence à une métrologie des sciences de la nature. Ce cadre épistémologique pragmatique-réaliste renvoie la nécessité de renégocier en permanence les catégories et mesures, c'est-à-dire de comprendre la démarche psychométrique comme une pratique relative à des objets que j'ai considérés comme « mous ». Cela veut dire rompre avec une forme d'expertise annoncée des mesures psychométriques, et ainsi considérer la mesure dans une démarche de dialogue avec la personne mesurée, donc dans une démarche de transparence autant par l'instrument utilisé que par l'usage de la mesure. C'est à dire concrètement pratiquer la mesure, via des modèles avec variables latentes dans une dynamique autre que la dynamique actuelle de normalisation et de

prétention à user de modèles valides, non vérifiables par les personnes, car soi-disant scientifiquement éprouvés.

3.2 Un pragmatisme usuel ?

Toute l'approche développée précédemment se positionne relativement au cadre du pragmatisme-réaliste, et donc en opposition aux courants dominants : l'empirisme-réaliste ou le constructivisme. Cependant de façon concrète le pragmatisme comme pratique concrète de la mesure semble en réalité usuelle. Cette pratique est discutée à travers « la validité » d'un test dans la littérature. Et en quelque sorte, toute ma discussion n'est qu'une discussion sur la « validité » d'une mesure.

3.2.1 Le pragmatisme dans la littérature

L'ouvrage collectif dirigé par Lissitz (2010) apporte une vision élargie des positions différenciées sur la validité des mesures effectuées en psychologie. Juhel & Gilles (2011) et Markus & Borsboom (2013a) reviennent aussi de façon synthétique sur ce concept de validité. Même si Borsboom a initialement défini la validité d'une façon différente (Borsboom et al., 2004), les développements du livre qu'il a co-rédigé avec Markus me laisse considérer que sa position n'est pas si éloignée des autres auteurs in fine. L'objectif ici n'est pas de revenir sur ce concept de « validité » car les synthèses citées ci-dessus suffisent.

Une forme de consensus se dégage sur la validité car, avec des nuances, la plupart des auteurs cadrent la validité dans une démarche théorique (théorie substantive) et une démarche pratique. Le curseur entre ces deux dimensions (théorie vs praticité) peut différencier entre auteurs, mais il semble y avoir une convergence sur la prise en compte nécessaire de ces deux aspects. En effet, la validation pour la plupart des auteurs passe par une boucle théorie-empirie, sans se focaliser sur la nature des outils statistiques (relativement peu puissants avec des variables latentes), ni à l'inverse sur la solidité apparente d'une théorie psychologique, car cette théorie est toujours enlisée dans un cadre social et historique et avec des effets d'inertie et de conformisme (Sijtsma, 2012b). De même le rapport de Juhel & Gilles (2011) sur la validité des tests psychométriques, bien que ne discutant que des deux courants dominants (empirisme-réaliste et constructivisme), positionne la pratique des tests dans une démarche pragmatique : toute la validité s'inscrit selon le contexte et objectifs définis par l'usage du test. Allant même plus loin, et en quelque sorte rejoignant la renégociation permanente

nécessaire que j'ai développée, Juhel et Gilles considèrent qu'aujourd'hui « *La validité d'une mesure prend donc, dans ces développements, un caractère relatif, devient une propriété évolutive, et la notion de validation, comme processus continu, remplace celle de validité* » (p. 13). La validité est pour ces auteurs d'une façon convergente aujourd'hui en phase avec Messick « *la validité est « un jugement évaluatif, intégré, du degré avec lequel les preuves empiriques, et les rationnels théoriques, plaident en faveur de l'adéquation et de la pertinence des inférences et des actions basées sur les scores* » (1989, p. 13). Cette définition met l'accent sur plusieurs points essentiels, qui sont acceptés aujourd'hui comme des principes de base de la validation (Kane, 2006). Elle réaffirme que la question de la validité est bien celle de l'interprétation et de l'utilisation des scores, ce qui implique de spécifier l'interprétation avant d'évaluer sa validité. Elle souligne que la validation de la mesure d'un construit théorique oblige à spécifier une théorie et la manière dont le construit est mesuré, ainsi qu'à évaluer empiriquement les prédictions de la théorie, ce qui nécessite des analyses extensives. Elle sous-entend enfin que l'interprétation des scores repose sur la mise en concurrence de plusieurs hypothèses et donc que d'autres interprétations des scores sont envisageables. ». Nous retrouvons : 1/ « *la validité est « un jugement évaluatif* » basé sur la théorie et les résultats empiriques : donc bien une démarche subjective d'objectivation ; 2/ « *l'interprétation et de l'utilisation des scores, ce qui implique de spécifier l'interprétation avant d'évaluer sa validité.* » : c'est donc bien une validité inscrite dans une pratique (enjeux et contexte) ; 3/ « *oblige à spécifier une théorie et la manière dont le construit est mesuré* » : on a bien une justification de la mesure basée sur un raisonnement subjectif (et non un cadre formel de la validation d'une mesure) ; 4/ « *d'autres interprétations des scores sont envisageables* » : c'est donc bien une « objectivité pour nous », inscrite dans une pratique, sans prétendre avoir « la vérité ».

Ainsi, la littérature se positionne en quelque sorte sur un terrain pragmatique. Je n'ai pas comptabilisé le nombre d'occurrences du mot « pragmatique » dans la littérature, mais il apparaît très fréquemment. Un auteur se réfère ouvertement au pragmatisme : Kane. Dans un long article (73 pages) il revient sur la validité des modèles avec variables latentes (Kane, 2013a) et l'historique de « la validité ». Dans sa réponse aux articles critiques, il affirme dans le titre de la démarche pragmatique (« *Validation as a Pragmatic, Scientific Activity* », Kane, 2013b). Pour lui « validité » intègre « *évaluation de l'interprétation et usage* » et use de « *interpretive arguments* » ou de « *interpretation/use argument* ». La validité n'est pas intrinsèquement propre au test, mais aussi à son usage. De même Messick, cité par Kane, écrit : « *an integrated evaluative judgment of the degree to which empirical evidence and*

theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. » Cependant pour Kane, il faut conserver une théorie forte sous-jacente car toutes les données peuvent potentiellement valider toute théorie. Kane (2013a) pose que le test est là pour prendre une décision, or c'est là que l'on glisse d'un outil d'aide à un outil réifié en vérité. Kane est prudent et dit que le score « présume » d'une performance, et c'est comme cela qu'il faut le prendre (« *presumptive in the sense that they can establish a presumption in favor of the conclusion but do not establish it definitively* » p. 11)

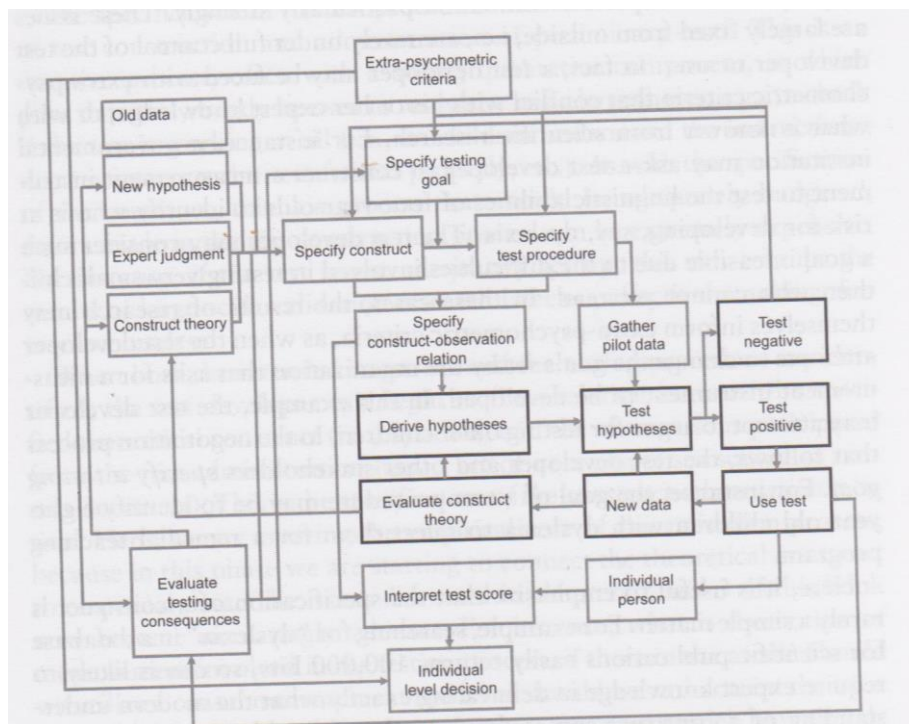
Mais Kane définit son pragmatisme (comme il le dit avec un « p » minuscule) plutôt comme une action pratique et non en référence au cadre épistémologique du pragmatisme, donc sans les soubassements définis précédemment : la nature ontologique des propriétés mentales mesurées et le cadre épistémologique pragmatiste nécessaire. On reste dans un cadre de pratique sociale d'un objet (Kane, 2013a), instrumentalisme déjà évoqué précédemment. Il faut donc, pour lui, avoir une démarche critique de la procédure. Tout son texte pointe la nécessité d'avoir une démarche avec un regard critique, critique de la réification, critique de la généralisation à partir d'un test sur une personne. La base est la corrélation, et on peut construire aisément des outils de décision absurde.

3.2.2 Le goldilocks de Markus et Borsboom

Je fais un focus sur le livre de Markus & Borsboom (2013a) car ce dernier couvre un large spectre de tout ce dont j'ai discuté. Positionné clairement dans un empirisme-réaliste (« *Natural kinds are types that are discovered rather than invented* » p. 237), Markus et Borsboom proposent en fin d'ouvrage en quelque sorte une forme de synthèse de la procédure de validation en tension entre théorie et pratique, avec laquelle je peux être en accord en grande partie. Très clairement, ces auteurs acceptent la validité d'un test dans une relation théorique et pratique, sans recours à un cadre formel unique, qui renvoie à une démarche pragmatique-réaliste ; d'ailleurs Markus et Borsboom usent du mot « démarche pragmatique » à plusieurs reprises pour justifier leur approche. La validation est ainsi un processus de discussion permanent, reconnaissant que la réalité en soi, donc la validation par un cadre formel, n'a pas de sens en psychologie. Ils précisent que la démarche est de réviser la théorie (la démarche empirique fait modifier les concepts). On retrouve en fait dans la pratique qu'ils proposent une démarche très proche du pragmatisme. Leur procédure (*goldilocks*, Figure 15) d'interprétation des tests est une approche balançant entre le réalisme nécessaire et la

subjectivité incontournable sur les objets psychologiques. Ils différencient ainsi la validité empirique et la validité épistémique d'un test, ce dernier étant d'un « niveau supérieur » et devant s'appuyer sur des justifications (théoriques) de « l'acteur » de la recherche : « *Because the epistemic claim involve an actor [...] the claim is to some extent relative to human standards and cognitions* » (p. 281), et en appelle au “consensus” comme les pragmatistes pour valider un cadre théorique. On retrouve cette démarche pragmatique chez Markus et Borsboom lorsqu'ils différencient les tests selon l'objectif et l'usage, donc “la pratique” (au sens étroit) est importante pour eux. Il y a chez Markus et Borsboom une volonté de faire rentrer le test dans le cadre formel du réalisme empirique, alors que tout tend à montrer que tout est « relatif » (construction, interprétation, ...) dans un test, tout en étant basé sur des réalités. En conclusion, toute la démarche opérationnelle proposée par Markus et Borsboom pourrait se rapprocher d'un pragmatisme-réaliste... mais qui nécessiterait de considérer la « réalité » des propriétés mentales avec ce même regard affirmé.

FIGURE 15 : PROCESSUS DE LA CONSTRUCTION ET DE LA VALIDATION D'UN TEST (GOLDILOCKS)



Tiré de Markus & Borsboom (2013a, p. 275)

3.2.3 Conclusion

En conclusion, de façon affirmée ou non, la littérature renvoie de fait à une démarche pragmatique avec un petit « p ». Car si la validation d'un test apparaît pragmatique, la démarche fondamentale ne l'est pas : la propriété mentale reste le plus souvent perçue comme une réalité fixe extérieure au chercheur. Ce n'est que la démarche pratique de la mesure qui devient, par nécessité, pragmatique, sans que les auteurs perçoivent la contradiction entre leur cadre épistémologique et leur cadre pratique, ni la contradiction entre l'hypothèse d'un objet réel extérieur à nos pratiques sociales et le besoin permanent de renégocier ces concepts. Ce qui est remarquable par exemple c'est le recours fréquent à l'abduction (Borsboom, Haig, ...) sans que cette démarche ne renvoie à une affirmation d'une démarche pragmatique, alors que l'abduction est consubstantiellement reliée à une vision de la science comme praxis et non pas comme une démarche simplement logique pour approcher « le vrai ».

Je considère qu'il faut se positionner en contradiction de la position « réaliste » dominante et de la position « constructiviste » et affirmer une approche pragmatique-réaliste non seulement sur la pratique de la mesure, mais aussi sur la pratique de la catégorisation de la propriété mentale. Sur la base d'une telle approche la propriété mentale conserve une objectivité, mais objectivité négociée dans un cadre pratique. Je le redis le pragmatisme auquel je me réfère n'est pas le pragmatisme avec un « p » comme Kane ou Messick car ce pragmatisme avec « p » est en fait un pragmatisme sur la base de l'empirisme-réaliste. Je défends un pragmatisme plus fondamental qui ne voit pas la dimension pratique uniquement dans l'objectif de la mesure, mais aussi sur l'objet mesuré, dans la pratique sociale de catégorisation. Ce qui justifie ce pragmatisme épistémologique c'est la nature des propriétés mentales comme propriété émergente de personnes nécessairement inscrites comme acteurs sociaux.

Une propriété mentale doit être considérée comme une objectivité catégorisable parce que renvoyant à des manifestations homogènes pour une communauté considérée. Ceci fait poser la question de la mesure de façon différente de Michell, Vautier, Markus, Borsboom, ... La méthodologie doit correspondre au cadre épistémologique, c'est-à-dire appréhender la mesure d'une façon pragmatique. Une mesure d'une propriété mentale demande relativement au pragmatisme (en tant que cadre épistémologique) de considérer la propriété mentale comme négociée selon le contexte et l'enjeu, demande de considérer la catégorisation relativement à son « unidimensionnalité » et « l'indépendance locale » des manifestations » (c'est donc la renégociation de la théorie critique qui pilote, construisant et déconstruisant ses catégories et

concepts), demande de valider si la propriété mentale est nominale, ordinale ou continue, c'est-à-dire s'il y a des sauts qualitatifs entre différents états possibles ou au contraire s'il y a une continuité « empirique » des différents états possibles. Toute cette démarche est inscrite dans une objectivité « pour nous », c'est-à-dire clairement en contradiction avec les travaux affirmant poser un cadre de mesure généralisable sans questionnement d'une propriété mentale. Mais donc aussi en critique des courants critiques de la mesure en psychologie se référant au cadre formel de la physique. Autant ceux qui « mesurent » sans questionner leurs pratiques, que ceux qui questionnent d'une façon critique la mesure, me semblent enlisés dans une approche empirique-réaliste en opposition à la réalité des propriétés mentales et le processus de connaissance lié.

Mais je ne considère pas ces deux courants opposés de manière identique. Le courant critique de la mesure en psychologie a l'intérêt indéniable de questionner une pratique sociale : le fait de mesurer à peu près tout et n'importe quoi sans questionner la validité de telles mesures.

3.3 *Soft measurement* ?

Caractériser la formalisation d'une mesure par une démarche pratique en rupture avec le cadre formel de la mesure renvoie à un courant qui développe un cadre du *soft measurement* (ou *weak measurement*). L'objectif n'est pas de développer in extenso ce courant qui fait l'objet de beaucoup de publications récentes, entre autres dans le journal *Measurement*, mais de poser ce qui me semble être des convergences et divergences avec ce que j'ai écrit précédemment, relativement à un courant qui, par ses préoccupations, semble proche de ce que j'ai proposé.

3.3.1 Les bases du *soft measurement*

Ce courant a initialement pour auteur central Finkelstein (Finkelstein, 2003, 2005; Grattan, Sun, & Khan, 2013; Khan & Finkelstein, 2013) qui discute du *soft measurement* pour des champs en dehors de la physique. La manière de poser les problèmes a des proximités avec ce que j'ai développé précédemment.

Les mesures (nominales, ordinales, d'intervalle ou de ratio) servent à s'extraire du flou linguistique : « *In weakly defined measurement, scales of measurement are classified in the literature according to the uniqueness conditions and their invariance under a class of transformations. Usually scales of measurement are classified as nominal, ordinal, interval*

and ratio. The wider the class of admissible transformations of a scale, the narrower is the class of meaningful statements in terms of the symbols of the scale [...] Description by natural language has often a degree of ambiguity, and vagueness. Description by measurement has generally a high degree of precision. Finally description by measurement is generally more concise than description by natural language. » (Khan & Finkelstein, 2013, p. 2938). Mais leur démarche se veut « objective », en rupture donc avec le cadre de Stevens.

Je cite longuement Finkelstein qui est la personne centrale de cette approche *soft measurement* et qui dans un article développe son cadre théorique (Finkelstein, 2005). Le cadre par lequel Finkelstein cherche à répondre au caractère mal-défini des concepts psychologiques, et donc à la difficulté d'appliquer le cadre formel de la physique aux mesures en psychologie, est : « *Measurement can be defined in the wide sense as a process of empirical, objective assignment of symbols to attributes of objects and events of the real world, in such a way as to represent them, or to describe them [...] The wide definition of measurement is often disputed by those who consider the paradigm of measurement in the physical sciences as normative, or, at least, require measurement to be a numerical representation in which reflects, at least, an order.* ». Finkelstein propose le terme de *weak measurement* pour donner ce cadre plus souple à une mesure respectant un processus objectif sans entrer dans le cadre formel de la physique : *'For this reason it is convenient to distinguish between strongly and weakly defined measurement. Strongly defined measurement is defined as a class of widely defined measurement, which follows the paradigm of the physical sciences. In particular, it has precisely defined empirical operations, representation by numbers and well-formed theories for broad domains of knowledge. Measurement that constitutes representation by symbols of properties of entities of the real world, based on an objective empirical process, but which lacks some, or all, of the above distinctive characteristics of strong measurement, may be termed weakly defined.* » (p. 268).

Khan & Finkelstein (2013, p. 2937) reviennent sur le *weak measurement* ou *soft measurement*. Ils reprennent un certain nombre d'aspects de la psychologie que j'ai développés. En psychologie, les mesures concernent des objets non-invariants, avec des concepts mal définis, sans pouvoir répondre au cadre formel de la mesure : « *Modelling and analysis by modelling of such soft systems pose significant challenges and it is made much more complicated by the fact that these systems involve human action, perception, feeling, decisions and the like. They can thus not be described by a system of invariant relations. The main difficulties for modelling in soft measurement systems stem from the fact that soft measurement systems (a)*

are based on ill-defined concept of quality, (b) have significant uncertainty in the empirical relational system that it represents, (c) have symbolic relational system with limited relations defined on it, and (d) have no adequate theory relating the measurement to other measurements in the same domain. ». Pour ces auteurs il y a des problèmes spécifiques dans différents champs de la science que ne connaît pas la physique : « *In fact the whole area of widely-defined measurement, which is needed for the wide and diverse application of measurement, offers significant conceptual problems, compared with measurement in the physical sciences. These are in relation to: (i) experiments and observation – economic and biological systems in which it is not always possible for models to represent reality adequately, (ii) replicability – in relations with measurands: psychological and social sciences – humans, complex systems where it may not be possible to formulate adequate models to represent such systems, (iii) utility – value judgement, quality and organisational performance measures, (iv) reliability, validity, generalizability – measurement in the social and psychological sciences, and (v) verifiability – economic and accountancy measurements, educational measurements.* »

Finkelstein (Finkelstein, 2003, 2005; Khan & Finkelstein, 2013) propose un cadre formel mathématisé de la mesure en psychologie que je ne développe pas car : 1/ il est très long (une dizaine de pages) ; 2/ il ne fait que formaliser mathématiquement les problèmes ce que je viens d'énoncer sans proposer réellement un nouveau cadre formel de validation d'une mesure en psychologie.

3.3.2 La démarche actuelle du *soft measurement*

Le *soft measurement* développe un certain nombre de réflexions sur la mesure et permet à ce courant d'évoluer. Mari par exemple essaye de poser, par une formalisation plus étendue que celle de Finkelstein, les convergences entre le *soft measurement* et la mesure en physique. (Mari et al., 2012). L'objectif est de s'extraire de la "boîte noire" qui fait assigner des valeurs à des objets par une "évaluation" en psychologie : « *If the generic process of assigning a quantity value to a quantity of interest is termed "evaluation" What does it characterize measurement as a specific kind of evaluation?* » (p. 2017) « *Pointing out that measurement is an evaluation abstractly characterizes it according to a black-box model, as a process realizing a functional transformation of an input entity to an output entity, called, in this case, the measurement result.* » (p. 2108). « *measurement is a map between the empirical world and the symbolic world, aimed at associating symbols to the (still unspecified, for the sake of*

generality) empirical entity under measurement in order to describe it [...] measurement is an informative property evaluation » (p. 2108).

A partir de cette position, ces chercheurs essaient de caractériser dans une démarche plus objectivante que la simple représentation posée par Finkelstein, basée pour Mari sur une objectivité et une intersubjectivité ; donc les termes sont proches de ceux que j'ai employés, mais pas forcément avec le même sens. Mari (2013) pose « *Our proposal is that for an evaluation to be considered a measurement its results must convey: – information specific to the measurand and independent of any other property of the object or the surrounding environment, including the subject who is measuring: this is a requirement of objectivity, i.e., relatedness to the object; – information interpretable in the same way by different users in different places and times: this is a requirement of inter-subjectivity, i.e., unambiguous representation.* » (p. 2893). Ainsi une mesure requiert une objectivité dans le sens où elle doit donner une information sur l'objet mesuré ; de plus la mesure doit être valide intersubjectivement, c'est-à-dire répliquable : « *two general features expected for measurement results [...] 1) Information specific to the measurand [“i.e., the property intended to be measured] and, therefore, to a given property of the object under measurement [...] This corresponds to guaranteeing that measurement results actually provide information about the measurand and not of some other property [...] it is a requirement of objectivity. 2) Information interpretable in the same way by different users in different places and times [...] a requirement of intersubjectivity* » (p. 2109). L'intersubjectivité implique un processus de calibration. Ce que Mari (2013) résume : « *Our proposal is that for an evaluation to be considered a measurement its results must convey: – information specific to the measurand and independent of any other property of the object or the surrounding environment, including the subject who is measuring: this is a requirement of objectivity, i.e., relatedness to the object; – information interpretable in the same way by different users in different places and times: this is a requirement of inter-subjectivity, i.e., unambiguous representation.* » (p. 2893)

Mari et ses co-auteurs dans différents articles posent que cette approche est en phase avec les normes de la mesure développées en physique.

- L'intégration de l'erreur d'une mesure est définie en physique par des « mesures incertaines » comme l'expliquent Mari et al. (2012) : « *Anyway, even if measurement might be hindered by errors (e.g., using a measuring instrument which is no longer calibrated), an appropriate intersubjective communication of measurement results calls for a statement about the degree of belief that the measurer attributes to the*

provided information. On this basis, the position emerged that a measurement result should specify the amount of information it is aimed at providing and, thus, that it is not complete if the uncertainty of the stated measured property value is not included. Specifically, the VIM3 defines “measurement uncertainty” as a “nonnegative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. »

- La validité n'est pas binaire en “oui” ou “non”, mais demande un processus de validation intersubjective : Mari et al. (2012) « *In practice, neither objectivity nor intersubjectivity is Boolean (i.e., yes–no) features [...] The threshold over which the evaluation results are “sufficiently” objective and “sufficiently” intersubjective to be considered measurement results is set in reference to the expected use of the provided information [...] ISO 9000:2005 standard: Any measurement process must first consider the intended use of the experimentally obtained information, in order to satisfy requirements predefined by the final user of that information. »*
- Le caractère “incertain” doit répondre à une norme posée : Mari et al. (2012) « *In particular, expressing uncertainty as a suitable standard deviation, as recommended by the Guide to the expression of uncertainty in measurements (GUM), allows comparing the quantity of the available information to the minimum quantity assumed by design as needed for effectively supporting the decision making, which is called the target measurement uncertainty (“measurement uncertainty specified as an upper limit and decided on the basis of the intended use of measurement results,” according to the VIM3) [...] if targets of uncertainty thresholds cannot be met, measurement results cannot be used. » (p. 2110)*
- L'opération de mesurage renvoie à une procédure en deux étapes : 1/ Modélisation ; 2/ opérationnalisation : Mari et al. (2012) « *The customary structure of a physical measuring system assumes that by means of measurement, an empirical entity (the measurand) is represented by an information entity (the measurand value including its uncertainty). This effect of bridging two different “worlds” results from a two-stage process: 1) a modeling stage, i.e., a set of conceptual activities needed to define the measurand and to model its relationship with the measurement result [...] 2) an operative stage, in which experimental activities are performed, so implementing the mapping between the empirical world and the world of symbols » (p. 2111)*

Au bilan, sans poser de cadre formel pour la psychologie, Mari et ses co-auteurs permettent de questionner le cadre formel de la mesure en physique qui est dans la pratique plus « pragmatique » qu'affirment par bien des auteurs, rejoignant en quelque sorte ce que j'ai écrit en m'appuyant sur Sherry (2011). Et donc, un peu comme le pose Sherry, la mesure en psychologie n'est peut-être pas si éloignée de la pratique concrète en physique et ces auteurs ont aussi recours au « pragmatisme » pour discuter de la mesure en psychologie.

3.3.3 Convergence et divergence avec le *soft measurement*

Il est délicat de situer l'approche que j'ai développée auparavant (valider subjectivement un isomorphisme entre une propriété et une échelle numérique par la validation subjective d'un continuum d'une propriété mentale) avec le *soft measurement* tel que défini par Finkelstein ou Mari. Mais je peux y voir une convergence : s'extraire du cadre formel qui ne peut s'appliquer en psychologie et essayer de donner un cadre formel pratique en résonance avec les objectifs de la psychologie.

Cependant, il me semble que la démarche pragmatique annoncée par ces auteurs rompt avec la démarche pragmatique que je propose. L'orientation dite « pragmatique » annoncée semble en effet considérer le pragmatisme comme répondant à une efficacité « pratique » et non dans un cadre épistémologique pragmatique tel que je l'ai posé ; notamment la nécessité de renégociation d'une mesure me semble disparaître dans leur formalisme. Ce courant me semble vouloir formaliser un cadre de mesure pour représenter une « réalité » sans questionner son ontologie. En effet, toute la discussion de ces auteurs sur les concepts psychologiques mal définis me laisse considérer que la démarche de ces auteurs sur la catégorisation des propriétés psychologique n'est pas en phase avec une démarche pragmatique de renégociation.

Le processus d'objectivation proposé par ces auteurs tend donc à essayer de se rapprocher de la démarche formelle en physique. Le cadre formel de la mesure en physique demande une calibration comme le pose Wright (1968) cité par Engelhard (2008) « *First, the calibration of measuring instruments must be independent of those objects that happen to be used for calibration. Second, the measurement of objects must be independent of the instrument that happens to be used for the measuring.* » (p. 87). Or la psychologie ne semble pas pouvoir fonctionner en deux étapes, Wilson, Mari, Maul, & Irribarra (2015) citent Duncan: « *Let us note the clear separation between the two tasks: first, to scale the statements, or to infer their location on a linear attitude continuum defined in terms of the polar contrast of "strongly*

affirmative” with “strongly negative”; and, second, to place respondents on that same attitude continuum insofar as their locations on it can be inferred from the selection of items they choose to endorse. Not only are the two problems logically distinct and operationally separated, but also the former – the scaling of statements – precedes and is presupposed by the latter [...] logically distinct and operationally separated processes [...] By contrast, psychometric or “test theory” methods of attitude measurement either attempt to accomplish the two tasks simultaneously and jointly or even to bypass the scaling of items altogether. ».

Wilson et al. (2015) considèrent qu'en réalité la psychologie procède de la même façon, en deux étapes, mais que cela rend le score normé dépendant d'une norme arbitraire qui n'est pas sans poser de problème (l'échantillon de référence), comme pour le QI par exemple. Cependant ces travaux sont trop récents pour conclure sur ce point. En conséquence Wilson (2013) et Wilson et al. (2015) discutent d'une convergence potentielle entre les mesures en psychologie et en physique sur la base de ce processus : définir l'instrument de mesure et la calibration, puis mesurer.

Si l'on peut définir la mesure en psychologie comme du *soft measurement* dans le sens où la démarche est plus « soft » qu'en physique (sur la calibration et l'unité de mesure essentiellement), je pense qu'il faut, à l'inverse du courant *soft measurement*, que la psychologie affirme son indépendance épistémologique relativement aux canons de la physique. Les objets mesurés répondent à des natures très différentes, comme le disent Hacking ou Quéré (que j'ai cités précédemment), ce ne sont pas des catégories « naturelles » (au sens de catégories fixes) qu'appréhende la psychologie, il faut donc une démarche en rupture avec de « vrais objets naturels ». La psychologie oblige à négocier la mesure dans une démarche pragmatique au sens où l'objectivation est inscrite dans une renégociation pratique des catégories utilisées. La mesure doit donc s'inscrire dans une telle démarche pratique en rupture avec le formalisme physique. Discuter comme le fait le courant *soft measurement* de la mesure en psychologie relativement à la physique est pertinent, mais l'orientation posée cherchant à faire correspondre la mesure en psychologie à la mesure en physique me paraît une erreur.

3.4 Conclusion

Je retrouve au final cette tension de bien des auteurs de valider la mesure en psychologie en marquant ses différences avec la physique, mais sans rompre avec le cadre empiriste-réaliste

de la physique ; entraînant ces auteurs à essayer de se justifier en référence au cadre de la physique qu'ils affirment pourtant ne pas pouvoir suivre. L'argument central devient dès lors : la physique aussi ne respecte pas son cadre formel... ce qui ne justifie en rien la mesure en psychologie, car le cadre expérimental en physique est absolument différent du cadre expérimental du cadre en psychologie, les objets investis sont de nature différents. Et la physique peut empiriquement valider la mesurabilité de ses objets... que ne peut faire la psychologie dans le cadre de l'empirisme-réaliste.

Seul un décalage épistémologique, affirmant que les objets de la psychologie sont des réalités objectives intersubjectivées, permet de s'extraire à mon avis de cette impasse en affirmant un cadre pragmatique-réaliste, non seulement dans la procédure de mesure, mais aussi de catégorisation des objets mesurés.

Partie 2 : REFLEXION METHODOLOGIQUE SUR LA MESURE
FORMATIVE D'UNE PROPRIETE MENTALE

Préambule à la seconde partie

Il peut paraître surprenant de voir ce sujet abordé suite à la première partie. Je me dois donc de justifier ce choix car, comme je l'introduirai par la suite, la première partie renvoie au renoncement de prétendre mesurer une propriété mentale de façon formative. Cependant j'ai décidé de conserver pour cette thèse cette seconde partie pour différentes raisons. Mais avant tout j'informe le lecteur que, par la suite, j'utiliserai du terme « construit » sans trop de précaution, comme la littérature académique l'utilise et comme j'ai considéré dans la première partie que nous pouvions le faire. Comme expliqué dans la première partie, le construit représente un objet qui est à la fois « l'objet réel » et à la fois « le concept », et donc l'usage de ce terme me facilitera le développement sur les mesures formatives sans devoir préciser si nous sommes au niveau de « la réalité » ou au niveau d'un « concept ».

La raison première de présenter ce travail sur les construits formatifs est qu'à l'origine de cette thèse je me sentais confronté à ces deux problèmes longuement discutés dans la littérature académique sur les variables latentes : 1/ A-t-on le droit d'utiliser des modèles avec variables latentes ? ; 2/ Si oui, a-t-on le droit d'utiliser des modèles « formatifs » ? J'ai donc mené cette double discussion parallèlement durant ma thèse. La première partie renvoyant le fruit de ma réflexion sur la première question ; la seconde répondant à mes interrogations méthodologiques. D'un point de vue ontologique, je considère aujourd'hui les construits formatifs comme ne pouvant prétendre représenter une propriété mentale (j'expliquerai cela dans mon introduction). Mais cette conclusion est le fruit du développement de ma première partie qui n'est arrivée à son terme qu'après avoir finalisé le travail effectué en parallèle sur les mesures formatives. Je m'étais en effet positionné de prime abord comme toute la littérature dans le cadre de l'empirisme-réaliste pour ma réflexion sur les modèles formatifs.

La première raison de rédiger cette partie est qu'il aurait été frustrant de ne pas rendre compte de tout un travail important effectué au cours de ma thèse. Mais je peux justifier cette partie en considérant que des auteurs peuvent être critiques de mon positionnement pragmatiste et affirmer eux travailler dans le cadre du réalisme-empirique, et donc prétendre pouvoir représenter une propriété mentale formativement. Donc, en prenant ce cadre du réalisme-empirique (que je considère aujourd'hui comme erroné), la question posée par la littérature est : un auteur peut-il user de construits formatifs sans risque ? Les construits formatifs sont en effet être critiqués sur le plan ontologique, mais aussi sur le plan méthodologique : les résultats seraient biaisés. C'est cette discussion d'ordre méthodologique que je traite ici.

Une seconde raison d'introduire ce travail sur les modèles formatifs répond à la nécessité d'illustrer à la suite de ma première partie les procédures de validation d'un modèle avec variables latentes par des simulations Monte Carlo, et la complexité associée pour extraire du sens opérationnalisable. En effet, comme expliqué précédemment, seules de telles simulations permettent de valider la robustesse de certains modèles. Or toute la discussion sur les modèles formatifs jusqu'à l'article de Bainter & Bollen (2014) ne s'était basée que sur quelques exemples... qui se révèlent porteurs potentiels d'erreurs d'interprétation comme je le discuterai lorsque je reprendrai les résultats des exemples proposés par la littérature (Franke, Preacher, & Rigdon, 2008; Howell, Breivik, & Wilcox, 2007b; G. Kim, Shin, & Grover, 2010; Wilcox, Howell, & Breivik, 2008). Il me paraissait incroyable que la littérature se serve depuis 2007 des trois exemples proposés par Howell, Breivik et Wilcox pour rejeter les modèles formatifs sans discuter que : 1/ trois exemples n'infirmant pas une méthode générale ; 2/ les critères retenus sont largement critiquables ; 3/ les données que ces auteurs ont utilisé ne pouvaient qu'apporter des résultats biaisés. Je considérais à l'amorce de ce travail qu'il fallait tout reprendre sur le problème des modèles formatifs, en posant une méthode fiable par voie de simulations. Mais la pratique des simulations Monte Carlo montre que ces simulations elles-mêmes doivent être discutées car il y a aussi dans de telles simulations des « choix » et « heuristiques ».

Une troisième raison renvoie à la nécessité de « publier » ce travail. Il est basé sur des simulations et a fait l'objet de tentatives de publications sans succès jusqu'à aujourd'hui, avec des refus de revues qui me semblent discutables sur leurs fondements. *Psychological Methods* a refusé de soumettre ce travail car il ne rentre pas dans la ligne éditoriale de la revue, or toute la discussion a été amorcée par un article Howell et al. (2007b) paru dans *Psychological Methods*. *Mis Quarterly* a refusé le papier car « il n'apportait rien de nouveau », je laisserai les lecteurs juge de cet avis après avoir lu la seconde partie de cette thèse mais le reviewer unique était Henseler, or notre travail proposé à *Mis Quarterly* intégrait des simulations sur PLS-SEM (l'approche *Partial Least Square*) et montrait les failles de cette modélisation pour les modèles formatifs... modèles PLS-SEM dont Henseler est un des hérauts. *Structural Equation Modelling* l'a refusé car PLS-SEM n'est pas opérant... alors que le travail soumis à cette revue avait été expurgé de toute discussion sur PLS-SEM et ne discutait que de CB-SEM. Une discussion avec le rédacteur en chef de cette revue, Marcoulides, m'a laissé très perplexe sur la lecture réelle de ce travail par les reviewers (sincèrement, je n'ai toujours pas compris comment un reviewer pouvait argumenter de façon critique un travail à partir de PLS-SEM lorsque celui-ci ne discute que de CB-SEM... ni comment Marcoulides pouvait

justifier cela). *Journal of Business Research* l'a refusé ensuite après trois tours. Un des reviewers visiblement n'avait qu'une connaissance lointaine de la discussion ; mais le second a clairement compris le travail et apporté des questions pertinentes (sur la standardisation des coefficients et l'usage des scores, que je développerai par la suite) et visiblement nos arguments ont convaincu ce reviewer... pour in fine le refuser car notre travail n'apportait aucun complément relativement à un court article publié dans *Measurement* (Guyon & Tensaout, 2015); donc de nouveau sans critique sur le fond. Il m'est donc apparu important de « publier » dans ma thèse l'ensemble des résultats obtenus... en attendant qu'une potentielle revue accepte ce travail (l'article est soumis à *Measurement* mais à l'heure où ces lignes sont écrites cette revue n'a toujours pas trouvé, deux mois après la soumission, de reviewers... et donc l'article est toujours en attente d'être soumis). L'ironie veut que juste après le refus de *Mis Quarterly* (qui prétendait que « tout avait été dit » ; c'était aussi un argument de *Structural Equation Modelling*) Bainter et Bollen ont publié un article en novembre 2014 dans *Measurement* sur la base de simulations relançant la discussion sur les modèles formatifs. J'avais envoyé à Bollen un projet d'article basé sur des simulations en décembre 2013 pour avoir son avis, et je n'ai eu aucune réponse... et c'est la seule fois où Bollen n'a pas répondu à un de mes mails. L'article de Bainter et Bollen n'est absolument pas un plagiat du travail envoyé un an avant... mais je suis en droit de penser qu'il y a une forte inspiration sur la méthodologie de simulations. J'ai donc repris suite à l'article de Bainter et Bollen l'ensemble des simulations que j'avais déjà effectuées pour apporter de nouvelles réponses à Bainter et Bollen, mais aussi à l'ensemble des articles ayant répondu à Bainter et Bollen (notamment Howell, 2014; Rhemtulla, Bork, & Borsboom, 2015), mais aussi pour apporter des compléments à notre travail (Guyon & Tensaout, 2015). C'est ce travail de simulations actualisé qui est reproduit dans cette thèse... en attendant qu'une éventuelle revue accepte de le publier. Il y eut en conséquence plusieurs phases de simulations avec quelques modifications de paramètres, ce qui explique par la suite parfois la non-homogénéité des paramètres utilisés. Mais cela ne modifie en rien les conclusions.

Ce sont donc pour toutes ces raisons que je présente l'ensemble de ce travail de réflexion et simulations. Mouloud Tensaout est co-participant de ce travail. Si j'ai la paternité de toutes les simulations et analyses de celles-ci dans notre travail collaboratif, je considère n'avoir aucune exclusivité sur l'ensemble de nos travaux car mes simulations ont été faites pour répondre aux questions que nous nous posons tous les deux, et de la manière de trouver des réponses efficaces. Donc si le code R (et MPlus) a été fait par moi-même, il est difficile d'extraire mon propre travail vs celui de Mouloud. Néanmoins je crois être en droit de les

présenter dans un travail personnel car j'ai l'accord de Mouloud pour ce faire. Je conserverai le « je » afin que Mouloud puisse conserver son libre arbitre sur mes interprétations et analyses.

J'informe enfin que ce travail est différent du court article publié dans *Measurement* (Guyon & Tensaout, 2015). Cet article a été écrit en une semaine (afin de pouvoir paraître dans le numéro suivant la publication de l'article de Bainter et Bollen) et ne discute que de quelques problèmes sans généraliser ni poser le cadre de simulations comme je le présenterai dans ce travail. Mais surtout nous avons fait une erreur grossière dans cet article : générer des construits réflectifs de second ordre avec seulement deux latentes de premier ordre (un tel modèle n'est pas théoriquement estimable). Les logiciels utilisés (MPlus et le package *sem* de R) ont donné des résultats et dans la précipitation nous n'avons pas vérifié in extenso tous les listings de sorties... ni les reviewers relever cette erreur. Cette mésaventure est assez pitoyable pour des personnes critiquant justement les erreurs usuelles dans la littérature académique (notamment notre article pointe une erreur des résultats de Bainter et Bollen que j'expliquerai par la suite)... mais l'auto-conviction est parfois tellement forte que même des auteurs s'estimant aguerris aux erreurs usuelles en statistique se laissent entraîner par leur volonté de prouver (certainement ce qui a dû arriver à Bollen pour l'erreur que je pointerai sur leur article par la suite). Ce qui est rassurant c'est de voir par exemple J. Gould faire la même autocritique (Gould, 1997).

4 Introduction aux mesures formatives

La discussion précédente (partie 1) portait sur la validité scientifique des modèles avec variables latentes. Nous n'avons pas abordé la distinction entre mesure formative et mesure réflexive d'une propriété mentale. Si nous le faisons dans cette seconde partie c'est pour essayer d'apporter des réponses pratiques à tout un débat sur les mesures formatives. La procédure pour valider empiriquement des modèles avec variables réfléchies semble faire consensus (excepté Rossiter comme je l'ai développé lorsque j'ai discuté des échelles de mesure). Par contre, les modèles avec variables latentes mesurées formativement renvoient à des questions sans réponse actuellement. Cela semble être un des sujets les plus discutés sur les modèles avec variables latentes, et où la communauté a le plus besoin de clarifier le débat car aucun consensus ne semble émerger. Ce travail empirique essaye d'apporter des réponses partielles à partir de simulations, toujours dans la perspective de discuter de la validité des modèles avec variables latentes. Avant de discuter de ces résultats, nous revenons brièvement sur les construits formatifs.

Je ne développerai pas les fondamentaux des approches CB-SEM et PLS-SEM et renvoie aux manuels qui discutent de façon exhaustive de ces modèles (manuels très complets... excepté pour les modèles formatifs) : Bollen (1989) et Kline (2010) pour CB-SEM et Hair et al. (2013) pour PLS-SEM. L'objectif des simulations proposées par la suite est un éclaircissement sur la possibilité d'intégrer une latente formative dans un modèle structurel, donc l'efficacité de l'estimation d'un tel modèle de mesure. Les simulations seront essentiellement axées sur CB-SEM qui aujourd'hui est considéré par la communauté comme répondant au cadre empiriste-réaliste, même si je développerai des simulations PLS-SEM. PLS-SEM est en effet considéré généralement comme inscrit dans une démarche qui peut s'apparenter à une démarche « constructiviste » et non « empiriste-réaliste » de par la formalisation mathématique du modèle de mesure. Mais ce point est lui-même un sujet de discussion controversé (voir par exemple Rigdon, 2012). PLS-SEM est en conséquence très discuté et il n'est pas dans l'objectif de mon travail de discuter de tous les aspects de cette approche qui renvoie à une littérature abondante et très récente.

4.1 La nature discutable d'une mesure formative

Les hypothèses associées à ces mesures formatives renvoient à des problèmes ontologiques et méthodologiques. Toutes ces discussions posent in fine la légitimité d'utiliser les mesures formatives (Edwards, 2011; Hardin & Marcoulides, 2011). Comme le souligne Diamantopoulos (2011), l'introduction de mesures formatives est « *considérablement* » plus compliquée à gérer sur le plan méthodologique que les modèles n'intégrant que des mesures réflexives. Il n'est pas alors étonnant de voir les experts reconnus se diviser entre les tenants du modèle de mesure formatif (Bollen, 2007; Diamantopoulos, 2011; Jarvis et al., 2003; MacKenzie, Podsakoff, & Podsakoff, 2011) et les opposants (Bagozzi, 2011a; Borsboom, 2015; Coltman, Devinney, Midgley, & Venaik, 2008; Edwards, 2011; Franke et al., 2008; Howell, Breivik, & Wilcox, 2007a; G. Kim et al., 2010; Markus, 2014; Wilcox et al., 2008). Je reviens dans cette partie sur la discussion académique et la critique de cette discussion.

4.1.1 Les mesures formatives récemment introduites

Contrairement aux mesures réflexives définies comme les manifestations d'un construit (Figure 16), les indicateurs formatifs sont vus comme les causes de la genèse du construit (Figure 17) (Bollen & Lennox, 1991; Edwards J.R. & Bagozzi R.P., 2000; Claes Fornell & Bookstein, 1982; MacCallum & Browne, 1993).

FIGURE 16 : UN MODELE DE MESURE FORMATIF D'UN CONSTRUIT DE PREMIER ORDRE

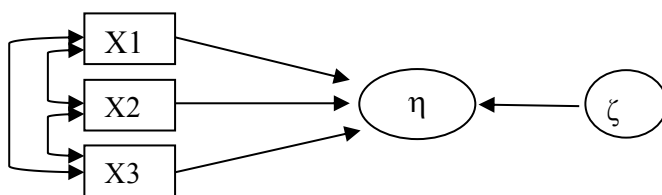
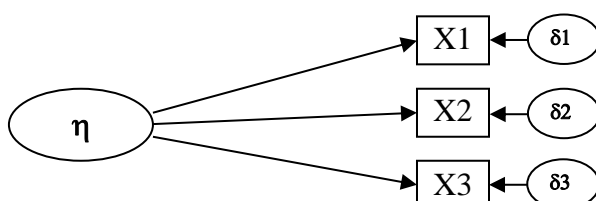


FIGURE 17 : UN MODELE DE MESURE REFLECTIF D'UN CONSTRUIT DE PREMIER ORDRE



La littérature académique a jusque récemment considéré comme seul recevable les construits mesurés réflexivement : Borsboom et al. (2003) « *In conclusion, the standard model in psychological measurement is a reflective model that specifies that the latent variable is more fundamental than the item responses. » ; Borsboom et al. (2004) : « *For instance, although the present validity concept can be applied directly to reflective latent variable models used in psychological measurement, it seems that formative models (Bollen & Lennox, 1991; Edwards & Bagozzi, 2000) do not allow for such application. In such models, the observed indicators are not considered to be causally affected by the latent variable but, rather, to cause such a latent variable. In this case, it is difficult to see how these observed indicators could be conceptualized as measures of the attribute in question because the arrows between the attribute and the observations run in the opposite direction. ».* Ce postulat était devenu une routine dans la pratique et la nature du construit (formative, réflexive) n'était pas discutée, il était considéré comme réflexif par définition. Le résultat de cette position est qu'un grand nombre de construits publiés, supposés réflexifs, ne le seraient pas en réalité (Bollen & Lennox, 1991; Jarvis et al., 2003; Rossiter, 2002). Depuis les articles de Diamantopoulos & Winklhofer (2001); Edwards J.R. & Bagozzi R.P. (2000); Jarvis et al. (2003), la production scientifique comporte donc nombre de travaux utilisant des mesures formatives. Mais, en conséquence, depuis les articles de Diamantopoulos & Winklhofer, (2001) et Jarvis et al. (2003), la discussion sur les modèles de mesures formatifs dans la littérature académique est prolifique et soulève de nouveaux débats. *Psychological Methods* (2007), *Journal of Business Research* (2008) et *MIS Quarterly* (2011), *Measurement* (2014) ont par exemple fait des numéros spéciaux sur ces construits mesurés formativement.*

4.1.2 Comment se formalise un construit formatif

Nous rappelons de façon très synthétique les travaux de Jarvis et al. (2003) et de Petter, Straub, & Rai (2007) qui posent les fondamentaux d'une mesure formative d'un construit : 1/ avec un construit mesuré formativement, le sens de causalité va des indicateurs au construit ; 2/ les indicateurs formatifs ne sont pas interchangeable ; 3/ les indicateurs ne partagent pas forcément un contenu identique ou similaire ; 4/ par contre, les indicateurs doivent couvrir l'exhaustivité du sens conceptuel du construit, donc supprimer un indicateur peut altérer le domaine conceptuel du construit ; 5/ la relation entre les indicateurs et le construit est linéaire ; 6/ le construit est unidimensionnel ; 7/ les indicateurs formatifs ne comportent pas

de terme d'erreur. Nous constatons donc que ces hypothèses sur les relations épistémiques entre le construit et ses indicateurs formatifs contraignent fortement le modèle de mesure. Mais ce qui complique encore les choses est que la plupart d'entre elles sont difficilement testables et ne peuvent être posées que par une argumentation théorique (le sens de la causalité, la linéarité, l'exhaustivité, et l'absence du terme d'erreur). Je reviendrai sur ces problèmes.

4.1.3 La nature ontologique des construits formatifs du point de vue empiriste-réaliste

Les mesures formatives renvoient en premier lieu aux questions conceptuelles traitant du statut ontologique d'un tel construit (Bagozzi, 2007, 2011b; Borsboom et al., 2003, 2004; Diamantopoulos, Riefler, & Roth, 2008; Diamantopoulos & Siguaw, 2006; Edwards J.R. & Bagozzi R.P., 2000; Howell et al., 2007a, 2007b; Markus, 2014; Markus & Borsboom, 2013b; Rhemtulla et al., 2015; Rossiter, 2002). Je mène de prime abord cette discussion par le prisme du réalisme-empirique car toute la littérature se base sur ce cadre pour en discuter.

Une mesure réflexive peut s'associer aisément à une démarche empiriste-réaliste car nous observons les manifestations d'une entité non-observable. Le point de vue réaliste pour une mesure formative est plus complexe à défendre pour un grand nombre d'auteurs. Avec une mesure formative, nous n'observons pas des manifestations d'une entité réelle (ou supposée l'être), nous hypothésons qu'une entité inobservable est construite par des indicateurs.

Est-ce un modèle de « mesure » ? Bagozzi (2011b) a fait un article de synthèse et il considère que le sens théorique d'un construit formatif n'est pas d'une nature différente qu'un construit réflexif, leur sens diverge uniquement au niveau empirique. En ce sens il adhère à ce que la littérature semble considérer aujourd'hui comme « une évidence », pour exemple Wilcox et al. (2008, p. 1220) « *Constructs themselves, posited under a realist philosophy of science as existing apart from their measurement, are neither formative nor reflective* ». Bagozzi (2011b) considère cependant que leur usage doit être réservé à de rares occasions : "*formative measurement and formative constructs have a place in research, but it is crucial to recognize that their applicability is restricted to a few narrowly defined models, unless one is willing to make a commitment to the ontology behind the approach and forgo the ontology and implications of the use of reflective measures exclusively* ». Markus & Borsboom (2013a) considèrent eux-mêmes que certaines mesures semblent devoir être faites formativement, prenant pour exemple la mesure de *l'exposition à la discrimination* dans leur livre... même si

pour eux cela ne peut être compris comme « une mesure » mais « *a formative measurement model represent a useful approximation rather a correctly specified causal structure under a process theory interpretation* » (p. 179).

Je comprends que des auteurs aient des difficultés de considérer « mesurer » un construit formativement, car le sens intuitif d'un modèle de mesure est le modèle réflectif. Mais cela n'interdit pas à priori dans le cadre du réalisme-empirique de représenter un construit formativement... et d'appeler in fine un tel modèle « une mesure » car les indicateurs formatifs permettraient de donner une estimation de la valeur du construit considéré. Donc la vraie question n'est à mon sens non pas sur la qualité de « mesure » d'un construit formatif, mais sur la nature ontologique d'un tel construit : peut-on encore considérer mesurer un objet réel à partir d'indicateurs formatifs ? Ou n'est-ce pas plutôt une démarche « constructiviste » ?

La démarche empiriste-réaliste est basée sur la considération qu'il y a des objets inobservables que nous essayons d'inférer. L'argument phare de Borsboom (entre autres) est que toute théorie sur une entité provient de manifestations observées, sinon elle ne serait pas testable : si une entité n'a pas de manifestations, cela rend impossible la démarche hypothético-déductive pour valider une théorie puisque nous n'en observons jamais les manifestations directes permettant de tester la théorie. Donc pour lui cela fait glisser clairement la démarche du réalisme vers le constructivisme. Cette argumentation souffre, comme déjà expliqué en introduction de ma première partie, du fait que la validité réflective se fait par un mélange de considérations théoriques et de résultats statistiques. Je le redis, aucun article basé sur l'empirisme-réaliste ne propose de cadre de validation d'une mesure au niveau de la propriété mentale (comme je l'ai proposé dans ma première partie), mais base la validité sur une discussion théorique et des résultats empiriques (validité statistique du modèle). La falsifiabilité d'un modèle réflectif ne paraît dès lors pas moins problématique que pour un modèle formatif, comme je l'ai discuté dans la partie 1. De plus, d'un point de vue « empiriste-réaliste » rien n'interdit de conceptualiser un objet mental d'une façon formative, considérant que cet objet mental existe mais qu'il n'est pas observable par des manifestations propres, mais uniquement appréhendable par des indicateurs provoquant la propriété mentale. Widaman (2014) considère par exemple qu'une propriété mentale mesurée formativement peut être considérée comme une « émergence » de ses causes (indicateurs) : « *so that the latent variable represents an emergent variable that arises due to the effects of its indicators.* » (p. 165). Je considère que le recours à l'émergence ici est un artifice explicatif car on ne précise absolument pas ce qu'est cette émergence en dehors d'une boîte noire explicative ;

mais fondamentalement dans le cadre usuel de l'empirisme-réaliste, ce qu'écrit Widaman est acceptable.

L'argument clé de la discussion actuelle sur la nature d'un construit formatif est sur le domaine unique potentiel (Lee, Cadogan, & Chamberlain, 2013; Markus, 2014). Bollen (voir entre autres Bollen, 2011 et Bollen & Bauldry, 2011) a beaucoup écrit sur cette nature unique de la mesure, différenciant les « *causal indicators* » (qui vérifient le domaine unique par l'introduction d'une erreur de mesure) et « *composite indicators* » (qui sont des index synthétiques de différents indicateurs). Mais comme Markus le rappelle (Markus, 2014) le domaine conceptuel unique est très compliqué à valider car comme les indicateurs doivent être peu corrélés entre eux, ils sont nomologiquement liés à des réseaux causaux à priori différents.

En conclusion, d'un point de vue empiriste-réaliste, les construits formatifs peuvent s'envisager... mais avec beaucoup de prudence. En les considérant comme des « approximations » (Markus & Borsboom, 2013a). En effet, le modèle formatif ne peut être considéré que comme un « proxi » de la mesure souhaitée... mais comme le sont aussi les mesures réflectives (voire Rigdon, 2012). Donc on peut très bien d'un point de vue empiriste-réaliste argumenter théoriquement pour justifier un modèle de mesure formatif, et poser le modèle de mesure comme un proxi utile, en discutant que de toute façon les modèles de mesure réflectifs sont tout autant des proxis utiles. Du point de vue empiriste-réaliste, il ne semble que rien ne puisse interdire le recours à des mesures formatives sur le plan ontologique... mais avec prudence (comme le dit Bagozzi, 2011b)).

4.1.4 Les mesures formatives d'un point de vue pragmatique-réaliste

Je ne vais pas reprendre tout ce que j'ai développé dans ma première partie. Je rappelle la position que je défends : une propriété mentale est un objet réel mais à appréhender comme une objectivité-intersubjectivée par ses manifestations.

4.1.4.1 Le non-sens d'une propriété mesurée formativement

Nous ne cherchons pas à mesurer un objet physique/biologique du cerveau, mais un objet réel en tant que propriété émergente, qui existe de par les interactions sociales, et qui ne peut donc être considérée comme réelle qu'à partir de ses manifestations. En conséquence, seuls des propriétés mentales ayant des manifestations « existent ». Je rappelle que cette position, qui peut sembler proche de ce que dit Borsboom, est en réalité différente de cet auteur car

Borsboom se positionne dans le cadre de l’empirisme-réaliste sans jamais clarifier la nature de « réalité » des propriétés mentales (je renvoie à ma première partie).

Ma position pragmatiste-réaliste ne veut pas dire qu’il ne puisse pas y avoir de facteurs exogènes (indicateurs formatifs) qui influencent des propriétés mentales, mais ces facteurs exogènes ne peuvent être considérés comme pouvant construire (mesurer) une propriété mentale. Si cette propriété mentale renvoie des manifestations observables, nous nous retrouvons clairement dans une schématisation MIMIC³⁰ (Figure 18). Pour un tel construit, les manifestations réflexives suffisent à la mesure, les indicateurs formatifs étant des causes exogènes... et nous retrouvons le schéma réflexif. Donc, clairement, un construit formatif n’a d’intérêt à être opérationnalisé que lorsque l’entité mesurée formativement n’a pas de manifestations visibles directes, mais influe sur d’autres entités inobservables (Figure 19). Voir pour des développements plus complets Lee et al. (2013) par exemple.

FIGURE 18 : MODELE MIMIC

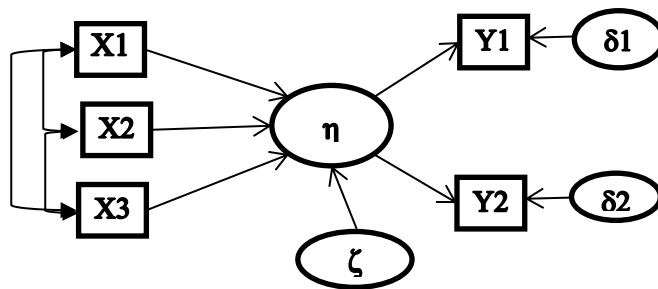
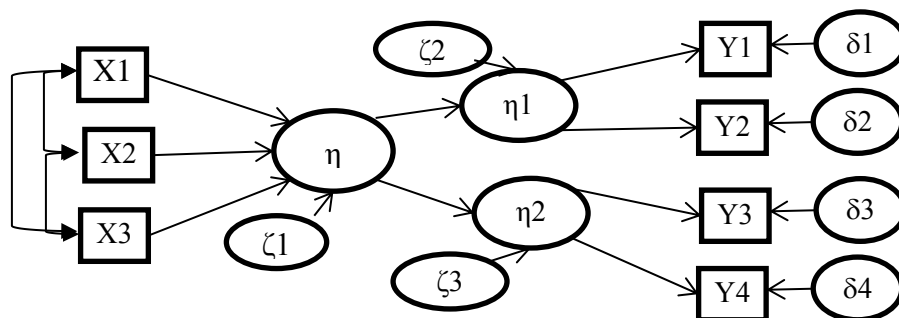


FIGURE 19 : VARIABLE LATENTE LIEE A DEUX AUTRES VARIABLES LATENTES



Donc la question centrale est la légitimité du modèle de la Figure 19. Or je considère à partir de ma première partie qu’un objet (représenté ici par η), n’ayant pas de manifestations, ne

³⁰ « Multi-Indicateurs (réflectifs) –Multi-Causes (indicateurs formatifs) »

peut être considéré comme une propriété mentale. Sans manifestation η n'a pas d'existence en tant que propriété mentale. Des indicateurs formatifs ne peuvent être vus que comme des facteurs exogènes à une propriété mentale dont nous pouvons observer des manifestations (Figure 18) et la Figure 19 est un non-sens ontologique.

Or toute la littérature discute en réalité du modèle posé Figure 19. Je crois que la discussion sur la nature du domaine unique d'une propriété mentale mesurée formativement renvoie à l'imprécision de ce que l'on cherche à mesurer dans le cadre de l'empirisme-réaliste. La discussion menée par Markus & Borsboom (2013a) par exemple se situe d'un point de vue empiriste-réaliste, donc un point de vue qui n'interdit pas à une propriété mentale d'exister même si elle n'a pas de manifestations propres (comme la Figure 19, ce que je conteste). Paradoxalement, toute la discussion critique des mesures formatives se fait sur l'argument qu'il faut partir des manifestations des propriétés mentales. Au final la discussion sur les mesures formatives renvoie à la discussion sur la nature de « réalité » des propriétés mentales, et la confusion/le manque de clarification de la littérature académique à ce sujet.

4.1.4.2 L'exemple des fonctions exécutives

Je reprends pour exemple l'article de Willoughby et al. (2014) sur les fonctions exécutives : « *The relatively recent term “executive functions” has been used to single out several higher order cognitive activities in an information processing model, such as strategic planning, impulse control, organized search, and flexibility of thought and action and self-monitoring of one’s behavior—activities that help maintain an appropriate mental set in order to achieve a future goal [...] Executive function—the ability to regulate behavior within a fluctuating and unpredictable environment—entails an integration of what the person wants to accomplish (hill), can do (skill), and directs energy toward (will) [...] Executive functions (EFs) are a set of general-purpose control mechanisms, often linked to the prefrontal cortex of the brain, that regulate the dynamics of human cognition and action* » (p. 25). Nous sommes bien ici dans la théorisation de processus mentaux complexes qui peuvent se voir comme basés sur 4 construits différents (A = executive function, B = inhibitory control, C = working memory/updating, D = shifting).

Ces auteurs proposent donc une mesure formative de la fonction exécutive... qui est bien une tentative de mesurer des processus complexes et non la mesure d'une propriété mentale en tant que manifestée et pouvant être considérée comme unidimensionnelle relativement à son cadre théorique. Cet article, comme déjà dit dans la partie 1, n'a qu'une vision floue de ce

qu'est réellement cette fonction exécutive en tant que « réalité », et comme déjà dit dans la partie 1 le recours à l'émergence sert à maquiller la nature « réelle » de cet objet car on prétend intégrer les processus complexes et l'émergence d'une propriété alors qu'il y a irréductibilité entre eux : « *To the extent that EF is an emergent property that is defined by interactions between neural systems, [...] To be clear, although neither formative nor reflective latent variables may adequately capture the variety or complexity of cognitive processes that support goal-directed behaviors, the suggestion here is that formative latent variables may provide a better approximation to modern theoretical accounts of EF than do reflective latent variables. Similar to our consideration of Jarvis et al.'s (2003) plausibility questions in the focus article, appealing to differences in the presumed ontological status of latent variables is a necessarily subjective process for which no definitive conclusions can be drawn* » (p. 177). Comme je l'ai déjà dit le passage des modules de base à l'émergence pose un problème dès que l'on part des modules de base car, alors, la propriété émergente est un artifice explicatif. Plus généralement, lorsque l'on lit un certain nombre d'articles usant de construits formatifs, on retrouve la confusion entre « processus mentaux » (système complexe) et « propriété mentale ». Les indicateurs formatifs sont en réalité présents pour « mesurer » des processus mentaux complexes et non une propriété mentale... processus complexes qui ne peuvent se mesurer (Ziegler & Vautier, 2014).

4.1.4.3 L'exemple de l'exposition à la discrimination

Si je reprends l'exemple donné par Markus & Borsboom (2013a), l'exposition à la discrimination, le fait de ne pas réussir à poser une mesure réflexive de cette propriété renvoie en fait à la « réalité » de cette propriété mentale. On retrouve pour cette mesure la méthode que je considère critiquable où le concept est posé sans questionner s'il répond à des manifestations réelles. Ayant posé un concept à priori (ici : l'exposition à la discrimination), on essaye de chercher les manifestations pouvant être considérées comme relevant de cette propriété théorisée... et parfois, nous n'y arrivons pas (parce que cette propriété mentale n'existe pas au sens de se manifester dans un cadre pratique). Comme on ne trouve pas de manifestations répondant aux impératifs d'une mesure réflexive (c'est-à-dire en réalité, on ne trouve pas de propriété mentale), on propose alors une mesure formative sous hypothèse que cette propriété mentale existe mais sans manifestations.

En disant cela je ne dis pas qu'il n'y pas « d'exposition à la discrimination », mais par contre peut être pas « d'exposition à la discrimination » au sens d'une propriété mentale.

L'exposition peut entraîner une propriété mentale (« sentiment de discrimination » par exemple), mais n'est pas en soi une propriété mentale. Le « sentiment de discrimination » peut par contre certainement être théorisé et mesuré en tant que propriété mentale. Peut-être que la mesure ne pourra pas se représenter sur une échelle numérique, car il faudrait creuser le sujet pour savoir s'il y a un potentiel continuum pour une telle propriété mentale ; ou au contraire ne faut-il pas considérer qu'il y a des seuils qualitatifs rendant impossible une échelle numérique. Par exemple le fait d'être d'une « population racisée » et donc potentiellement discriminée n'est certainement pas sur un continuum relativement à un homme blanc socialement installé ; il doit y avoir des sauts qualitatifs dans la mesure du « sentiment de discrimination ». De même, certainement, le ressenti de discrimination diffère selon que nous l'abordons via le genre ; via l'ethnie ; via les pratiques sexuelles catégorisées (homosexuel vs hétérosexuel par exemple) ; ou les catégories de sexes (transgenres par exemple). Bref le besoin de poser une mesure formative dans cet exemple de discrimination illustre à mon sens l'erreur du cadre théorique liée à cette mesure : par une mesure formative on essaye de mesurer via des indicateurs des processus complexes générant potentiellement une propriété. Mais si cette propriété mentale existe, elle se manifeste.

4.1.4.4 L'exemple du statut socio-économique

L'exemple le plus usuel pour la discussion sur les mesures formatives est le *Statut Socio-Economique* (SES) appréhendé par le revenu, l'éducation et l'habitat (c'est l'exemple que pratiquement tous les articles prennent pour discuter du fondement des mesures formatives). De nouveau ce statut socio-économique est une entité abstraite, un indicateur sociologique à l'image de ce que Desrosières (2008b) discute. Mais un tel concept n'en fait pas une propriété mentale existante chez les personnes. Il existe des mesures réflexives du statut social tel que vécu par les personnes, qui partent de manifestations à partir de la conscientisation de ce ressenti de statut, mais nous ne discutons alors plus du tout du même objet. Le SES mesuré formativement est un index utilisé en sociologie et ne peut prétendre représenter une réalité mentale définie par ses indicateurs formatifs. Peut-être que les indicateurs (le revenu, l'éducation et l'habitat) sont des facteurs susceptibles de générer des processus mentaux, et que ces processus mentaux génèrent in fine une propriété stable pour une personne (le ressenti du statut social). Mais c'est donc bien cette propriété mentale de « ressenti » qui nous intéresse en tant que devoir être mesurée. Cela n'interdit pas d'introduire des indicateurs exogènes (revenus, ...) pour essayer de comprendre des dynamiques psychologiques mais

sans prétendre par-là « mesurer » une propriété mentale car si l'objet mesuré a des manifestations, il faut partir de ses manifestations. S'il n'en a pas... c'est que l'on discute de processus complexes sans manifestations, donc nous ne discutons plus d'une propriété mentale.

4.1.4.5 Conclusion

A partir de la Figure 19, les indicateurs formatifs ne peuvent prétendre que mesurer des processus complexes qui génèreront d'autres processus complexes qui eux-mêmes génèreront une propriété émergente (qui se manifestera dans l'espace social). Autant η_1 et η_2 ont un sens comme représentants des propriétés mentales (en tant qu'émergence) ; autant η ne peut prétendre représenter une émergence mais juste un processus complexe. Or on ne mesure pas des processus complexes pouvant générer d'autres processus complexes, on cherche en psychologie à mesurer une propriété mentale qui doit donc se manifester pour être considérée comme « réelle » (je renvoie à ma discussion sur les *Network Analysis* qui prennent comme base, avec raison, exactement ce même regard).

En conclusion, il ne peut y avoir de mesure formative d'une propriété mentale car ces dernières sont des réalités sociales, et ne sont objectivables que par des manifestations. La démarche doit partir de ces manifestations et donc poser des mesures nécessairement réflexives.

4.1.5 Les problèmes méthodologiques

Même sans adopter le point de vue que je défends, une mesure formative renvoie à beaucoup de problèmes méthodologiques. Toute la littérature se positionne dans le cadre du réalisme-empirique acceptant la formalisation de la Figure 19. Si la littérature académique reconnaît aujourd'hui la légitimité de ces construits formatifs (Bagozzi, 2011b; Slaney et al., 2011), indépendamment de la légitimité ontologique d'un construit, un examen critique du processus de développement du modèle de mesure formatif est crucial.

La littérature académique discute notamment de la difficulté d'estimer un modèle formatif de façon fiable et donc de l'ambiguïté d'interprétation de ces mesures formatives. Ce problème a été particulièrement discuté dans le cadre de l'analyse de covariance (Bagozzi, 2007, 2011b; Bollen, 2007, 2011; Bollen & Davis, 2009a, 2009b; Cadogan & Lee, 2013; Cenfetelli & Bassellier, 2009; Diamantopoulos, 2010, 2011; Diamantopoulos et al., 2008; Diamantopoulos & Siguaw, 2006; Edwards, 2011; Finn & Wang, 2014; Franke et al., 2008; Hardin, Chang, &

Fuller, 2008; Hardin et al., 2008; Hardin & Marcoulides, 2011; Howell et al., 2007a; Jarvis et al., 2003; G. Kim et al., 2010; MacCallum & Browne, 1993; MacKenzie et al., 2011; Marakas, Johnson, & Clay, 2008; Wilcox et al., 2008), que nous associerons à l'algorithme le plus connu CB-SEM. L'approche non paramétrique *Partial Least Square Path Modelling* (PLS-SEM) est alors apparue comme une solution alternative car jugée moins contraignante et plus robuste pour des échantillons de faibles tailles et violant l'hypothèse de normalité (Gefen, Rigdon, & Straub, 2011; Ringle, Götz, Wetzels, & Wilson, 2009). Son succès est aussi facilité par l'existence de nombreux logiciels récemment développés (SMARTPLS, le module PLS-SEM de XLSTAT, PLS GRAPH, le package *plssem* ou *sempls* sous R, SPAD, ...). La discussion a donc légitimement dérivé dernièrement sur la pertinence de l'approche PLS-SEM pour les modèles structurels intégrant des mesures formatives (Bollen, 2011; Diamantopoulos, 2011; Goodhue, Lewis, & Thompson, 2012; G. Kim et al., 2010; Pelagatti, Fattore, & Vittadini, 2012; Ringle, Sarstedt, & Straub, 2012).

Il existe maintenant des travaux qui récapitulent l'ensemble des faiblesses associées au développement d'un modèle de mesure formatif et qui menacent sa validité (Diamantopoulos, 2011; Hardin & Marcoulides, 2011; G. Kim et al., 2010). Elles peuvent être regroupées en deux catégories : 1/ celles inhérentes aux hypothèses contraignantes pour le développement d'un modèle de mesure formatif, jugées comme irréalistes (p.e. absence d'erreur de mesure) ou difficiles à satisfaire (p.e. exhaustivité des mesures) (Edwards, 2011) ; 2/ celles qui portent sur l'instabilité des paramètres du modèle de mesure formatif qualifiée dans la littérature « d'interprétation confondante »³¹ (Howell et al., 2007b). Cette dernière désigne le cas où le sens nominal (théorique) d'un construit diverge du sens empirique (Howell et al., 2007a, 2007b).

Pour illustrer simplement les problèmes inhérents aux mesures formatives utilisées dans la littérature académique, nous reprenons le travail de Chin (2010). Cet auteur propose un modèle formatif de « *la ressource perçue* » inclus dans différents MES et un modèle réflexif ; le tout étant estimé par PLS-SEM. Le modèle réflexif est clairement posé et semble fiabilisé. Par contre, la mesure formative est introduite dans différents modèles et les résultats des estimations montrent que, les « poids » des indicateurs formatifs (coefficients de la régression multiple des indicateurs sur le construit estimé) et les « *loading* » (coefficients de la régression simple des indicateurs formatifs sur le construit estimé) changent fortement d'un modèle à un autre. Je donne un tableau comparant les liens entre les indicateurs et le construit

³¹ *Interpretational confounding*

« *la ressource perçue* » entre deux modèles différents proposés par Chin (2010), le premier « modèle de redondance » relie simplement le modèle de mesure formatif de la « ressource perçue » au même construit mesuré réflectivement ; le second est un modèle dit « global » liant « *la ressource perçue* » à d'autres construits (tous représentés avec des mesures réflectives). Au regard des valeurs données par Chin (Tableau 4), il semble délicat de considérer que nous avons affaire à un seul et même construit « *ressource perçue* » formé par les sept mêmes mesures, notamment au regard de la variable R10 (« *documentation* ») dont le poids passe de -0,022 à +0,4120.

TABLEAU 4 : POIDS ET LOADING (ENTRE PARENTHESES) POUR LE CONSTRUIT "RESSOURCES PERÇUES"

Mesures	Modèle de redondance (*)	Modèle final (**)
<i>Hardware software</i> (R5)	- 0,589 (0,930)	- 0,0024 (0,5832)
<i>Knowledge</i> (R6)	0,270 (0,814)	0,2284 (0,7290)
<i>Time</i> (R7)	0,132 (0,654)	0,5611 (0,8549)
<i>Financial ressource</i> (R8)	- 0,100 (0,735)	- 0,1015 (0,5597)
<i>Someone's help</i> (R9)	0,027 (0,566)	0,0698 (0,5941)
<i>Documentation</i> (R10)	- 0,022 (0,546)	0,4120 (0,7592)
<i>Data</i> (R11)	0,118 (0,602)	0,0874 (0,6847)

(*) Chin, 2010, fig 28.1 p. 683 (**)Chin, 2010, fig 28.2 p. 686

Cet exemple illustre ce qu'on appelle « l'interprétation confondante » sur lequel je reviendrai en détail. Cet exemple illustre les biais potentiels même avec PLS-SEM, pourtant considéré usuellement comme devant être utilisé avec des mesures formatives (voir par exemple l'article de synthèse et de recommandations de Gefen et al. (2011).

4.1.6 conclusion

Devant le débat sans consensus sur l'usage des modèles formatifs, Hardin & Marcoulides (2011) préconisaient d'autres travaux afin d'apporter à la communauté scientifique une vision claire de la pertinence des modèles de mesure formatifs. En effet, les hypothèses associées aux construits mesurés formativement sont jugées par de nombreux auteurs comme fragiles ou irréalistes et posent la question de la légitimité des mesures formatives à représenter un construit (Bagozzi, 2011b; Borsboom et al., 2003; Edwards, 2011; Howell et al., 2007b).

Je considère que la psychométrie doit avoir un cadre reconnu pragmatique-réaliste, et dans ce cadre les mesures formatives contredisent ce qu'est une propriété mentale car cette dernière n'existe que par des manifestations. De tel construits formatifs sont en conséquence clairement inscrits dans une démarche constructiviste, comme explicité précédemment. Cependant, par la suite, je vais travailler sous hypothèse « empiriste-réaliste », c'est-à-dire en considérant potentielle la mesure formative d'une propriété mentale. En effet, un chercheur peut bien considérer que mon point de vue est erroné... et donc justifier le choix d'une mesure formative par le recours au « réalisme » tel que posé par Wilcox et al. (2008) que j'ai cité précédemment (puisque ce point de vue est quasi-consensuel). Mais alors, même de ce point de vue, est-ce possible ? Il faut donc clarifier le débat sur l'efficacité des SEM à modéliser des mesures formatives dont les estimations soient fiables. Nous discuterons en substance par la suite de la principale objection aux mesures formatives : la dépendance des mesures formatives au modèle structurel spécifié. L'objectif de notre travail est de revenir sur le plan méthodologique sur les difficultés de développer un construit mesuré formativement. Puis, par simulations Monte Carlo, nous chercherons à définir le cadre pratique pour lequel une telle approche peut s'avérer efficace.

4.2 Contraintes et menaces sur les mesures formatives

Je discute des contraintes et menaces liées aux construits mesurés formativement avant de développer des simulations pour clarifier le débat.

4.2.1 Hypothèses formelles entre le construit et ses indicateurs formatifs

Le processus de développement du modèle de mesure formatif pour représenter un construit (Figure 19) repose sur l'examen d'un ensemble d'hypothèses formelles sur les liens entre le construit η et ses indicateurs X_i . Ces principes et hypothèses constituent la théorie auxiliaire qui établit la correspondance entre le construit et ses indicateurs. Je pars du postulat que la définition conceptuelle du construit peut être opérationnalisée avec des mesures formatives. Je vais présenter le processus de développement de ces mesures qui s'appuie sur une théorie auxiliaire (Bagozzi, 2007).

Afin de mieux comprendre les implications pratiques de ces hypothèses auxiliaires, je présente les deux principales spécifications du modèle formatif qui correspondent aux deux approches traditionnelles développées dans la littérature sur les MES, à savoir CB-SEM (CB-SEM) et PLS-SEM (PLS-SEM). Je rappelle les équations car elles impliquent un sens différent au construit avec CB-SEM ou avec PLS-SEM.

4.2.1.1 Le cadre d'analyse de CB-SEM

Dans le cadre de CB-SEM, le construit mesuré formativement est défini comme une variable latente composite η (VLC) qui est une combinaison linéaire de ses mesures X_i et d'un terme d'erreur :

$$\eta = \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \zeta \quad (6)$$

ou sous forme matricielle :

$$\eta = \Gamma \Xi + \zeta \quad (7)$$

Cette formalisation du modèle de mesure dans un cadre paramétrique suppose plusieurs hypothèses et contraintes importantes sur la distribution du terme d'erreur, et la variance du construit, à savoir :

$$E(\zeta) = 0, E(\zeta / X) = 0 \text{ et } \text{COV}(X_i, X_j) = \Phi$$

$$V(\eta) = \gamma_1^2 V(X_1) + \gamma_2^2 V(X_2) + \gamma_3^2 V(X_3) + 2 [\gamma_1 \gamma_2 \text{COV}(X_1, X_2) + \gamma_1 \gamma_3 \text{COV}(X_1, X_3) + \gamma_3 \gamma_2 \text{COV}(X_2, X_3)] + V(\zeta) \quad (8)$$

Mais en contrepartie, elle nous permet de tester diverses hypothèses rivales et ainsi de pouvoir juger de la validité (ou non) du modèle de mesure à l'aide de nombreux indices d'ajustement fournis (voir la partie validation).

A titre de comparaison, le modèle réflectif d'un construit η , Figure 17, s'écrit :

$$X_i = \lambda_i \eta + \delta_i \quad (9)$$

Avec $E(\delta/\eta) = 0$

$$V(X_i) = \lambda_i^2 V(\eta) + V(\delta_i) \quad (10)$$

Le terme d'erreur représente les autres facteurs non pris en compte car jugés peu importants pour altérer la signification du construit. CB-SEM bénéficie d'indices de fiabilité du modèle, néanmoins, il faut garder à l'esprit que l'interprétation de ces indices reste fragile. En effet, lorsque ces indices rejettent le modèle, il est vraisemblable que le modèle doit être légitimement rejeté. Par contre dans le cas d'indices satisfaisants, le modèle n'est pas forcément correct (comme nous l'illustrerons par la suite par des exemples). D'où le risque important d'accepter un modèle de mesure formatif mal spécifié en se basant uniquement sur ces indicateurs statistiques (RMSEA, GFI, ...).

4.2.1.2 Le cadre de la modélisation PLS-SEM

Nous rappelons les bases de PLS-SEM. Nous reprenons les écritures matricielles identiques à celles de CB-SEM avec les mêmes conventions d'écriture que Fornell & Cha (1994) plutôt que l'écriture analytique de Tenenhaus, Vinzi V.E., Chatelin Y.-M., & Lauro C. (2005), mais cela ne change rien à la compréhension de l'algorithme. Je renvoie à l'article de Tenenhaus et al. (2005) pour un développement exhaustif de PLS-SEM et à Riou, Guyon, & Falissard (2015) pour un développement plus illustré.

Soit le SEM standard (mesures réflectives, mais avec un modèle formatif cela ne change rien à l'algorithme, seul les estimations finales sont modifiées) :

$$\eta = B\eta + \Gamma\xi + \zeta \quad (11)$$

$$y = \Lambda_y \eta + \varepsilon \quad (12)$$

$$x = \Lambda_x \xi + \delta \quad (13)$$

L'objectif ici est de trouver des relations pondérées entre les variables latentes et les variables manifestes :

$$\eta = \pi_\eta y \quad (14)$$

$$\xi = \pi_{\xi} x \quad (15)$$

Nous illustrons cet algorithme sur un MES comportant deux construits. Notons par X_1 la matrice des mesures réflexives de la variable latente V_1^i , et X_2 la matrice des indicateurs formatifs de la variable latente V_2^i ; l'exposant "i" indique le nombre d'itérations de l'algorithme.

A l'étape zéro, la variable latente V_1^0 est la première composante normée de l'ACP. La variable latente V_2^0 est choisie arbitrairement à l'étape "0". A la première itération, nous construisons de nouvelles variables : V_1^1 est donnée par les corrélations entre le premier axe factoriel et la variable V_2^0 . La variable V_2^1 est estimée par la régression de la variable (arbitraire au premier pas) V_1^0 sur X_1 . Ceci donne la première estimation externe des variables latentes.

Soit :

$$V_1^1 = (X_1'X_1)^{-1} X_1'V_1^0 \quad (16)$$

$$V_2^1 = X_2X_2'V_2^0 \quad (17)$$

Suit une estimation, dite "interne", des relations entre les différentes variables latentes. Pour ce faire trois possibilités s'offrent, sachant que ces paramètres semblent très peu changer les résultats finaux (Tenenhaus et al., 2005) : structurel, centroïde ou factoriel.

De nouvelles variables V_1 et V_2 sont calculées : V_1^2 et V_2^2 . On détermine de nouvelles relations externes :

$$V_1^3 = (X_1'X_1)^{-1} X_1'V_1^2 \quad (18)$$

$$V_2^3 = X_2X_2'V_2^2 \quad (19)$$

L'algorithme itère jusqu'à convergence, qui est assurée pour des cas simples (deux variables latentes) et extrêmement probable pour des cas complexes (les cas de non stationnarité semblent rarissimes) (voir Hanafi, 2007; Henseler, 2010). Les relations entre variables latentes (modèle interne) sont les régressions linéaires entre les dernières estimations externes de celles-ci.

Dans le cadre de la modélisation PLS-SEM, le construit formatif est supposé déterminé entièrement par ses indicateurs de mesures et en conséquence aucune autre cause n'impacte sa variance (Cenfetelli & Bassellier, 2009; Tenenhaus et al., 2005). Contrairement à CB-SEM,

dans l'approche PLS-SEM le construit est défini comme une variable composite (VC) et n'incorpore pas l'erreur de mesure.

$$C = W_1 X_1 + W_2 X_2 + W_3 X_3 + W_4 X_4 \quad (20)$$

Cette hypothèse complique davantage la validation du modèle de mesure (Bollen, 2011; Diamantopoulos, 2011; Hardin & Marcoulides, 2011). Mais, d'un point de vue pratique, l'élimination du terme d'erreur est aussi commode dans la phase d'estimation du modèle formatif (absence de cadre paramétrique qui permet de relâcher bien des hypothèses contraignantes avec LIREL), et est à l'origine du succès de l'approche PLS-SEM (Ringle et al., 2012).

4.2.2 Identification du modèle de mesure formatif

L'estimation des paramètres du modèle de mesure spécifié nécessite au préalable son identification. Je présente maintenant les étapes d'identification et d'estimation du modèle de mesure nécessaire pour examiner la validité de construit (validités prédictive et nomologique). Rappelons que cette phase empirique de la validation d'un modèle de mesure est au cœur des controverses récentes. Cette étape d'identification est effectuée uniquement dans le cadre de CB-SEM (Bollen & Davis, 2009a, 2009b). Par contre, avec CB-SEM et PLS-SEM, nous rencontrerons des soucis dans la phase d'estimation.

4.2.2.1 Dans le cadre de CB-SEM

Récemment, Bollen & Davis (2009a, 2009b) ont proposé un ensemble de règles à satisfaire pour identifier les paramètres d'un modèle de mesure formatif.

1. *La règle deux variables émises*

Cette condition nécessaire mais non suffisante impose l'ajout de deux variables endogènes à la variable latente η mesurée formativement. Concrètement, nous pouvons soit développer des mesures réflectives associées à la latente formative η , soit spécifier d'autres construits avec des mesures réflectives (Bollen & Davis, 2009a) (Figure 18 et Figure 19), ou une combinaison des deux. Par exemple, le modèle illustré par la Figure 18 qualifié de MIMIC est identifié. Mais il faut s'assurer que théoriquement ces nouvelles variables ne sont liées qu'au travers de la variable latente formative. Cette hypothèse dite d'indépendance locale signifie que les deux variables Y1 et Y2 sont conditionnellement indépendantes par rapport à la

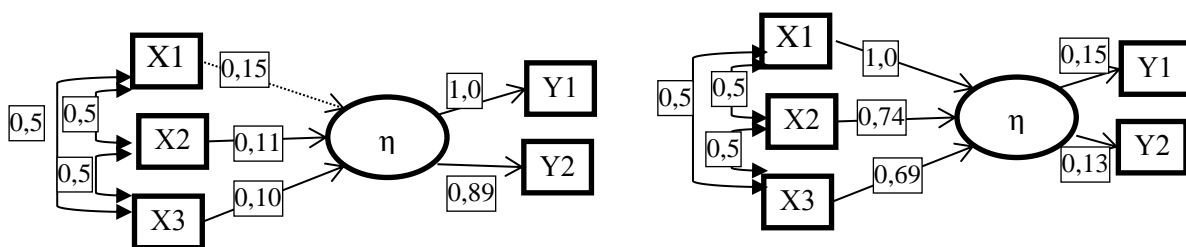
variable η , sinon le cadre mathématique obligatoire n'est plus respecté et les résultats seront biaisés. C'est une contrainte forte et difficilement testable dans la pratique. Nous reviendrons sur ce point. Enfin ce schéma d'identification induit des contraintes de proportionnalité (voir plus loin).

Notons qu'il n'est malheureusement pas rare de voir des articles spécifiant un modèle de mesure formatif sans que ce dernier ait des liaisons avec au moins deux autres mesures réflexives ou deux variables latentes endogènes avec des mesures réflexives. Théoriquement un tel modèle n'est pas identifiable par CB-SEM. Pour y parvenir, des contraintes supplémentaires spécifiées *à priori* sont nécessaires (en fixant une valeur nulle à certains paramètres ou à celle de l'erreur du construit). Mais une telle variable formative ainsi spécifiée (c'est-à-dire sans deux liaisons avec des variables endogènes) est intégrable dans n'importe quel modèle structurel car qu'elle soit ou non présente, cela ne change rien au modèle. C'est pourquoi on la qualifie de "variable fantôme" (Rindskopf, 1984). Mais en aucun cas on peut prétendre avoir mesuré ce construit, il est purement artificiel et sans aucune validité psychométrique (Bagozzi, 2007; Diamantopoulos & Siguaaw, 2006; MacCallum & Browne, 1993; MacKenzie et al., 2011).

2. La règle de normalisation des variables latentes

La norme d'une variable latente est arbitraire, on peut fixer à « 1 » l'un des paramètres liant le construit à l'une de ses mesures ou à un autre construit ; une autre option consiste à fixer à « 1 » la variance du construit ou la variance résiduelle de la variable latente. Mais selon le choix de calibration choisie, les résultats diffèrent. J'illustre ce problème par deux estimations issues d'un même jeu de données.

FIGURE 20 : A) MODELE ESTIME AVEC LE LIEN ENTRE ETA ET Y1 CONTRAINT ; B) MODELE ESTIME AVEC LE LIEN ENTRE X1 ET ETA CONTRAINT



Quel que soit le choix de la normalisation, les résultats standardisés donneront théoriquement les mêmes estimations. Cependant, selon les logiciels utilisés, les résultats peuvent diverger

en fonction des algorithmes intégrés (Franke et al., 2008). Il ne faut donc impérativement analyser que les résultats standardisés pour comparer les coefficients (Bollen, 2011). La contrainte est que seuls les logiciels Mplus et package LAVAAN donnent des intervalles de confiance de ces estimations standardisées, mais LAVAAN semble avoir des bugs pour estimer une latente formative³². Donc la procédure conseillée sera de vérifier par bloc (c'est-à-dire par variable latente) la significativité des coefficients non-standardisés, puis d'analyser les coefficients standardisés de tout le modèle (en ne retenant que les coefficients significatifs). Cependant les tests de significativités posent problèmes (pour des raisons algorithmiques) et deux normalisations différentes peuvent voir les tests (Student) donner des résultats divergents (alors que mathématiquement ils devraient être les mêmes); voir Diamantopoulos (2011) et Franke et al. (2008). Franke et al. (2008) conseillent de prendre comme norme la variance de la latente à 1 et de faire des tests *Likelihood Ratio* (LR) pour vérifier la significativité des paramètres. Concrètement on compare par un test du chi2 un modèle contraint et un modèle libre, paramètre par paramètre (il faudra dès lors contrôler par Bonferroni par exemple). Cette procédure est lourde mais semble impérative... et je n'ai jamais vu un seul article l'ayant effectuée.

4.2.2.2 Dans le cadre de l'approche PLS-SEM

J'ai souligné que cette méthode non paramétrique est moins contraignante pour estimer le modèle de mesure. Cette approche connaît ainsi un succès pour estimer un MES en raison de sa capacité à surmonter les difficultés rencontrées avec CB-SEM (identification, solution impropre, convergence, taille de l'échantillon, distribution des variables). Je ne discute pas de tous les aspects de PLS-SEM (et de ses extensions récentes), mais uniquement de sa capacité à modéliser un modèle formatif d'une variable latente (cf. Vinzi, Chin, Henseler, & Wang, 2010, pour un développement exhaustif).

1. Identification

Dans cette approche, l'étape de l'identification ne se pose pas. En effet, PLS-SEM possède la propriété de pouvoir estimer un modèle de mesure formatif d'une variable latente, même si celle-ci n'est pas reliée à des mesures réflectives ou à d'autres construits endogènes. Mais un tel modèle de mesure estimé isolément ne sera en fait que la première composante principale de ses indicateurs formatifs. Le résultat serait identique si nous spécifions un modèle de

³² J'ai essayé de poser ce problème aux développeurs de LAVAAN sans succès.

mesure réflectif ! De même, avec PLS-SEM, une variable latente peut n'être reliée qu'à une seule autre variable latente (formative ou réflective), mais là aussi nous retrouvons un cas particulier de ce que l'on nomme l'analyse multi-blocks (Tenenhaus & Hanafi, 2010). Donc PLS-SEM n'apporte réellement d'intérêts par rapport aux outils statistiques usuels que lorsque le modèle formatif est relié à au moins deux autres variables latentes ou mesures dans le modèle.

2. Contraintes sur les covariances

L'algorithme PLS-SEM impose *à priori* que les corrélations entre les mesures du modèle formatif estimé soient libres³³. Par conséquent, lorsque nous souhaitons, pour des raisons théoriques, contraindre certaines de ces corrélations (p.e. à 0), les estimations du modèle fournies par PLS-SEM sont alors biaisées (Cenfetelli & Bassellier, 2009).

3. La qualité d'ajustement

Le cadre non paramétrique de PLS-SEM ne fournit pas véritablement d'indices d'ajustement du modèle spécifié aux données ni de seuils pour le rejet ou non d'une hypothèse portant sur les propriétés souhaitées d'une mesure valide (Hair et al., 2013). Notons que Tenenhaus et al. (2005) ont discuté d'index globaux de qualité, comme le « *gof*³⁴ » (voir Ringle et al., 2012; Wetzels, Odekerken-Schröder, & van Oppen, 2009, pour des applications pratiques). Mais ces derniers n'ont pas de propriétés paramétriques connues, comme celles de l'approche du maximum de vraisemblance utilisée dans CB-SEM et aucun GOF ne semble potentiel aujourd'hui (Hair et al., 2013). Enfin, le recours à des indices d'ajustement non paramétriques³⁵ fournis par la méthode PLS-SEM ne permet pas de comparer des modèles rivaux.

4.2.3 Le processus de validation d'un modèle de mesure formatif

Le processus de validation diffère sensiblement de celui d'un modèle de mesure réflectif. Je discute à l'aide de simulations simples ce problème afin d'illustrer la fragilité de ce processus de validation.

³³ Ce n'est pas le cas avec CB-SEM

³⁴ Pour PLS-SEM, Wetzels, Odekerken-Schröder, & van Oppen (2009) proposent des seuils en fonction du GOF pour juger de la validité d'un modèle (rejet) : un seuil minimum de 0,25 est jugé acceptable.

³⁵ Il faut comparer ces indices souvent à des seuils *ad hoc* pour juger de la qualité d'ajustement du modèle aux données. C'est aussi le cas pour certains indices non paramétriques (mais pas tous) fournis par CB-SEM.

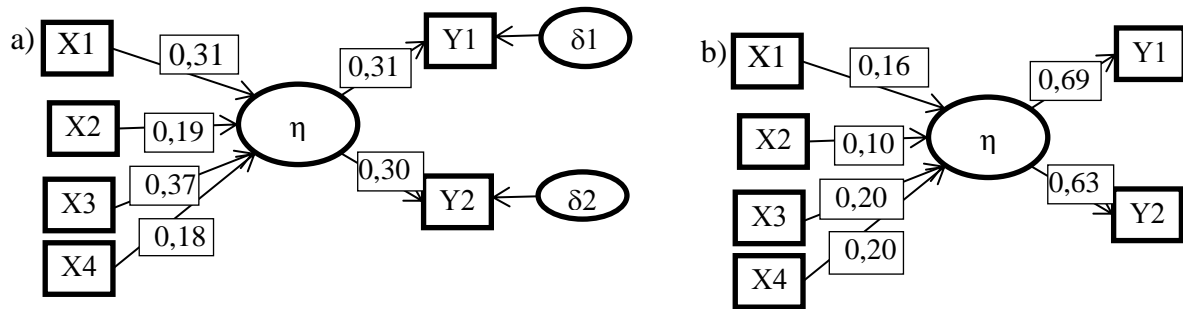
4.2.3.1 Description des simulations

J'ai souligné plus haut que chacune des deux modélisations (CB-SEM et PLS-SEM) repose sur un nombre d'hypothèses très contraignantes dont certaines sont difficiles à satisfaire (exhaustivité des indicateurs formatifs, absence d'erreur de mesure). Afin de faciliter la compréhension de cette discussion sur les fragilités des hypothèses qui sous-tendent un modèle de mesure formatif, j'ai construit des jeux des données simulées à partir d'un modèle de mesure théorique d'une latente formative avec 4 indicateurs formatifs indépendants et 2 mesures réflectives pour des besoins d'identification. L'ensemble des jeux de données sont simulées avec un échantillon de 500 individus par le logiciel R qui permet de travailler sur les mêmes données avec CB-SEM (package *sem*) et PLS-SEM (package *sempls*). Les estimations des intervalles de confiance des paramètres sont obtenues sur la base de 1000 répliques Bootstrap. Dans tous les modèles, pour éviter les problèmes de colinéarité mais aussi pour ne pas alourdir le schéma du modèle théorique, les covariances entre les indicateurs formatifs sont nulles sauf indication contraire.

Nous construisons de prime abord un modèle de mesure formatif valide. Le poids standardisé γ_i de chacune des mesures formatives est fixé à 0,25. Enfin, les coefficients entre la variable latente et les mesures réflectives sont de 0,30. Il va de soi que ces corrélations sont faibles pour des mesures réflectives valides, néanmoins ces dernières sont ajoutées uniquement pour des besoins d'identification du modèle de mesure formatif qui nous intéresse ici. Enfin ce modèle de mesure est théoriquement valide et respecte donc les contraintes de proportionnalités.

Les résultats des estimations de ce modèle avec CB-SEM et PLS-SEM sont donnés par la Figure 21. Il est à noter que tous les coefficients estimés sont significatifs. Un premier constat qu'il convient de rappeler est que les estimations obtenues par PLS-SEM et CB-SEM divergent sensiblement, PLS-SEM apportant un poids supérieur aux variables réflectives. Cette distorsion de PLS-SEM envers les variables manifestes (réflectives) est reconnue par la communauté usant de PLS-SEM (Hair et al., 2013). Proportionnellement les ξ_i jouent le même rôle avec CB-SEM et PLS-SEM, mais le praticien verrait à tort dans le modèle avec PLS-SEM un rôle plus important des variables réflectives.

FIGURE 21 : MODELE DE BASE A) ESTIME AVEC CB-SEM * ; B) ESTIME AVEC PLS-SEM **



* Chi 2 : 2,75 ; p value : 0,43 ; RMSEA :0,000

** gof=0,252

4.2.3.2 L'hypothèse de normalité des données

PLS-SEM est théoriquement libre de contraintes de normalité des données. Avec CB-SEM, le relâchement de la contrainte de normalité des distributions des données est sans graves conséquences sur les paramètres estimés des mesures réflectives si l'aplatissement et l'asymétrie ne sont pas trop éloignés de leurs valeurs théoriques. Mais ce n'est pas le cas pour des indicateurs formatifs : leurs distributions doivent être impérativement gaussiennes (Ringle et al., 2009). Il semble y avoir le même problème avec PLS-SEM (Ringle et al., 2009). Il faudra donc impérativement faire des tests de normalités des distributions des indicateurs formatifs X_i avant toute chose quel que soit l'approche (CB-SEM ou PLS-SEM).

4.2.3.3 L'exogénéité de l'erreur de la variable latente

Seul CB-SEM modélise le modèle avec un terme d'erreur associé à la variable latente formative. PLS-SEM n'intègre pas d'erreur de mesure de la latente. Dans le cadre de CB-SEM, (Diamantopoulos & Siguaw, 2006) ont longuement discuté le rôle et l'interprétation du terme d'erreur ζ (équation (6)) dans le processus de validation du modèle de mesure formatif. Il est maintenant admis que ce terme ne peut représenter que les autres indicateurs jugés secondaires sans impact réel sur la signification du construit, qui doivent être, comme dit précédemment, indépendants des indicateurs inclus dans le modèle (je développerai une discussion critique de cette erreur par la suite). Cette hypothèse d'exogénéité des erreurs est cruciale pour l'utilisation des méthodes de régression. Ainsi, la minimisation de ce terme d'erreur est le point clé pour juger de la validité du modèle spécifié. En conséquence une variance de ce terme d'erreur ζ faible et des indicateurs significatifs avec des signes attendus

sont nécessaires dans le processus de validation. Lorsque la variance est jugée importante³⁶, alors un réexamen des mesures retenues est fortement recommandé. Si en plus les signes et/ou des indicateurs sont non significatifs alors une nouvelle définition du contenu doit être entreprise. Enfin le cas d'une variance faible et des indicateurs non significatifs peuvent indiquer une forte colinéarité entre des indicateurs formatifs.

Il reste à déterminer les causes qui peuvent violer cette hypothèse d'exogénéité. La littérature économétrique cite trois sources : l'erreur de mesure, l'omission d'une variable importante et la simultanéité entre des variables³⁷. Nous avons souligné plus haut que l'estimation du SEM par les méthodes de régression exige que le terme d'erreur de la variable latente ne soit pas corrélé avec les mesures X_i retenues (hypothèse d'exogénéité). Cette condition implique concrètement que les facteurs influençant η omis dans le modèle doivent impérativement être indépendants des variables X_i introduites, sous peine d'avoir des résultats biaisés.

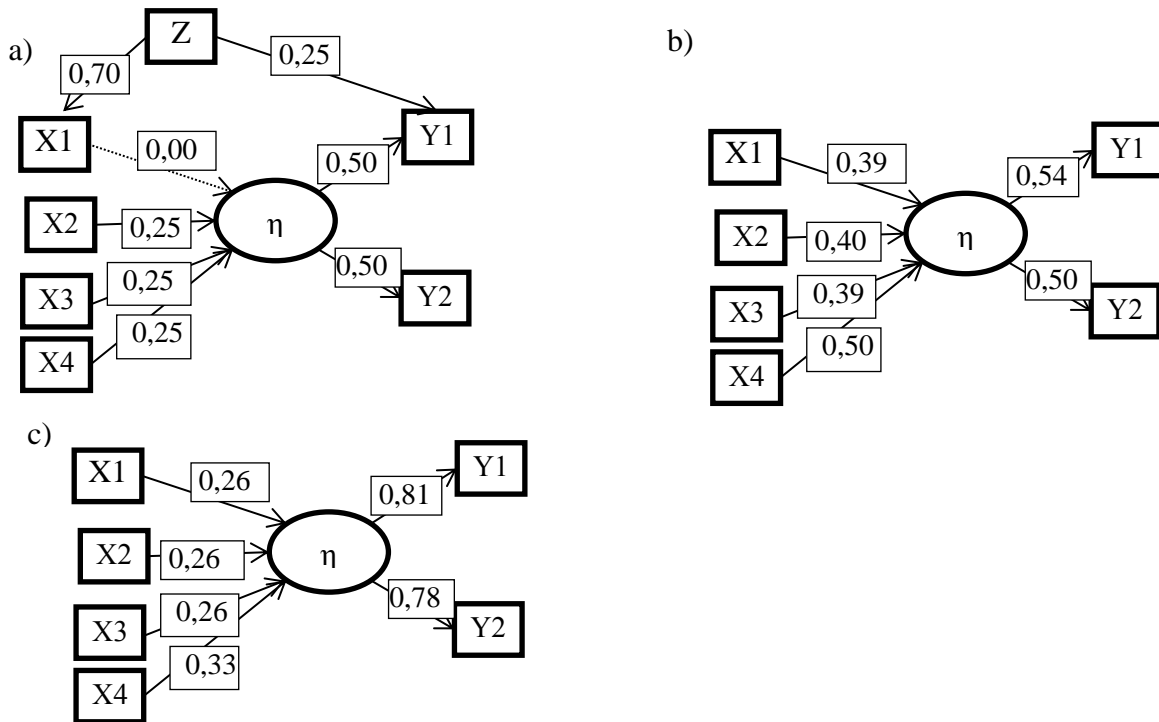
J'examine un exemple classique (qui consiste à spécifier une mesure formative X_1 , alors qu'elle ne l'est pas car liée à une autre mesure formative omise Z (Figure 22)). Supposons que le praticien ne retienne que X_1 , X_2 , X_3 et X_4 pour définir le construit η et n'intègre pas la variable Z dans son modèle, l'hypothèse d'exogénéité est violée et les résultats seront biaisés. En effet, ζ (erreur de la variable latente, cf. équation (6)) correspond à tous les autres facteurs influençant η non intégrés dans le modèle (dont Z). Théoriquement ζ doit être indépendant des facteurs intégrés. Or ici, ζ n'est pas indépendant de X_1 puisque Z influence X_1 et Z ³⁸ est intégré dans ζ . C'est ce que l'on appelle la "non exogénéité" de l'erreur, problème très fréquent souvent ignoré qui entraîne des conclusions fausses. Concrètement, X_1 apparaîtra influencer significativement η (Figure 22 b/ et c/) à tort (Figure 22 a/). Les indices d'ajustement ne permettent pas de déceler ce problème. Ainsi, ce problème d'endogénéité ne peut être discuté que théoriquement à partir de la définition du construit et des indicateurs retenus. Il est donc important d'avoir cette réflexion pour le développement de ces mesures car aucun test statistique ne permet de détecter une telle mauvaise spécification, à moins de développer plusieurs modèles rivaux.

³⁶D'après Diamantopoulos se référant aux travaux de Cohen (1978) pour juger de l'importance de l'ampleur de la taille de l'effet, une variance expliquée du terme de l'erreur ζ de plus 25% est jugée excessive.

³⁷ C'est le cas lorsque la variables dépendante, ici η et une variable explicative, ici X_i , sont déterminées simultanément (comme pour l'offre et de la demande d'un bien).

³⁸ La variable Z omise est qualifiée dans la littérature de « variable confondante».

FIGURE 22 : A) MODELE THEORIQUE ; B) MODELE MAL SPECIFIE ESTIME AVEC CB-SEM * ; C) MODELE MAL SPECIFIE ESTIME AVEC PLS-SEM **



* Chi 2 : 2,23 ; p value : 0,53 ; RMSEA : 0,008

** gof=0,47

En conséquence, la seule solution pour éviter ce problème est de justifier théoriquement que nos indicateurs formatifs sont des causes directes et exhaustives de la formative³⁹. Il reste le problème de l'absence de l'erreur de mesure. Nous examinerons ci-dessous ces deux points : exhaustivité des mesures et absence d'erreur de mesure.

4.2.3.4 Des mesures exhaustives

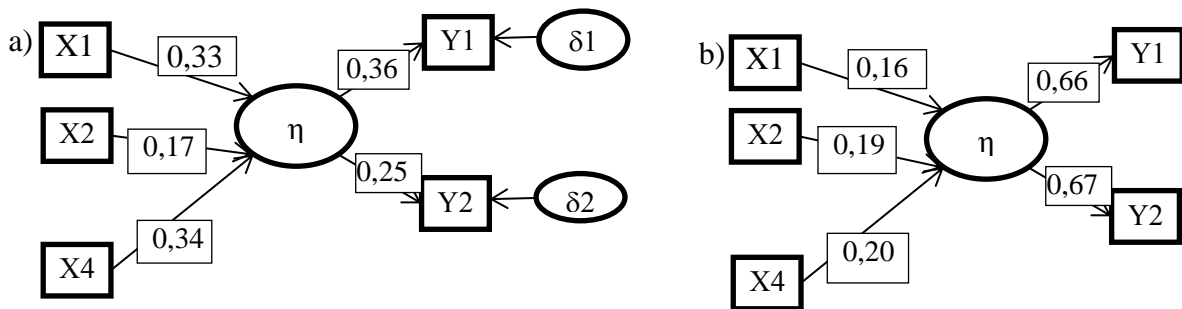
L'une des conditions essentielles pour la validité de contenu d'un modèle de mesure formatif est de s'assurer que l'ensemble des causes du construit d'intérêt sont identifiées. Par définition, l'omission de l'une d'entre elles (X_i) peut altérer la signification du construit postulée *à priori*. Selon Diamantopoulos (2011), l'omission d'une des causes importantes du construit peut conduire à altérer le contenu du construit et donc son interprétation. Néanmoins, dans le cadre de CB-SEM, cette définition est moins stricte en incluant un terme

³⁹ Nous reviendrons sur ce problème lorsque nous effectuerons des simulations pour apporter des outils empiriques de réponse.

d'erreur ζ qui représente d'autres causes jugées peu importantes (équation (6)). La seule condition exigée est que ce terme d'erreur ne soit pas corrélé avec les mesures X_i retenues (hypothèse d'exogénéité discutée précédemment). Pour illustrer ces propos, nous avons éliminé du modèle de base (Figure 21), l'indicateur formatif X3 (le plus influent, Figure 23). Nos variables X_i ne sont pas corrélées entre elles, donc les coefficients estimés ne seront pas beaucoup modifiés. Ce cadre de covariances nulles permet de ne pas considérer que l'information enlevée (ici X3) se transmet via les autres indicateurs conservés dans le modèle (X1, X2 et X4). Ce qui est remarquable ici (Figure 23) c'est que les indices d'ajustement des modèles estimés restent très satisfaisants (RMSEA=0 ; gof=0.25). Sur la base de ces résultats, un praticien jugerait que ce modèle de mesure est valide alors qu'il est incorrectement spécifié (la variable X3 est supprimée)⁴⁰.

Ainsi la validité de contenu d'un modèle de mesure formatif ne peut être examinée que sur la base de la définition conceptuelle du construit et le jugement par les pairs.

FIGURE 23 : MODELE SANS X1 A) ESTIME AVEC CB-SEM * ; B) ESTIME AVEC PLS-SEM **



* $Chi^2 : 0,580 ; p \text{ value} : 0,748 ; RMSEA : 0,000$

** $gof=0,250$

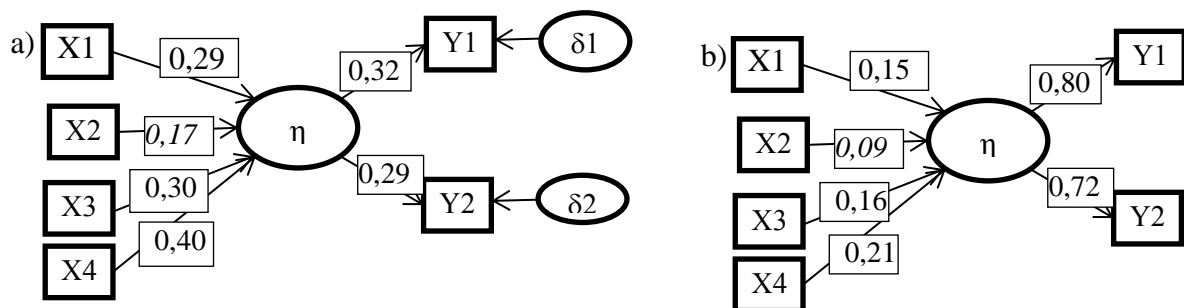
4.2.3.5 L'absence de l'erreur de mesure des indicateurs formatifs

Comme le montre l'équation (6), les indicateurs formatifs ne comportent pas d'erreurs de mesures. J'ai souligné plus haut, bien que cette hypothèse soit jugée peu réaliste dans la pratique notamment lorsque les mesures sont obtenues par un questionnaire (Edwards, 2011), cette hypothèse est maintenue pour les besoins d'estimation du modèle de mesure. Donc concrètement, les indicateurs formatifs sont présumés exacts, sans erreur possible alors que le

⁴⁰ Nous reviendrons sur ce problème et les moyens pratiques de vérifier l'exhaustivité des indicateurs formatifs

plus souvent ce n'est pas le cas. Par conséquent il faut s'attendre à des paramètres estimés biaisés. Afin d'illustrer l'importance de ce biais nous avons repris le modèle de base (Figure 21) et ajouté des erreurs aux mesures des indicateurs formatifs X_i (erreurs représentant 10% des variations des mesures). Nous avons ensuite estimé ce nouveau modèle de mesure (Figure 24). Nous constatons que les poids estimés des indicateurs diffèrent significativement par rapport à leurs valeurs initiales (Figure 21), la variable la plus importante apparaît être X4, et X2 a un coefficient non-significatif ce qui amènerait à retirer cette variable du modèle. Les indices d'ajustement sont toujours très satisfaisants.

FIGURE 24 : MODELE AVEC UNE ERREUR DE MESURE A) ESTIME AVEC CB-SEM * ; B) ESTIME AVEC PLS-SEM **



* $\chi^2 : 2,42 ; p \text{ value} : 0,489 ; RMSEA : 0,000$

** $gof=0,30$

Cet exemple illustre l'importance de cette hypothèse implicite d'absence d'erreur pour les indicateurs formatifs. Pour répondre à ce problème, Edwards (2011) préconise tout simplement une reformulation du modèle formatif dans laquelle chacune des mesures devient une manifestation (mesure réflexive) d'un construit de premier ordre. Pour Edwards (2011), le recours aux mesures réflexives permettrait de s'affranchir du problème d'interprétation. Cependant cette formulation souffre, au-delà d'un problème d'interprétation puisque la variable latente formative semble devenir une variable latente réflexive, de devoir imposer les erreurs des indicateurs, c'est à dire choisi arbitrairement par le praticien. Cette procédure proposée par Edwards a été très peu utilisée et demanderait d'autres investigations à partir d'exemples ou simulations. En effet, sur nos données, en entrant l'erreur réelle des indicateurs (puisque nous la connaissons ici exactement), les résultats sont surprenants puisque le coefficient de X1 devient significativement négatif. Bien évidemment la modélisation

proposée par Edwards est impossible avec PLS-SEM. Nous rediscuterons de ce problème par simulations pour clarifier les menaces de mesures formatives.

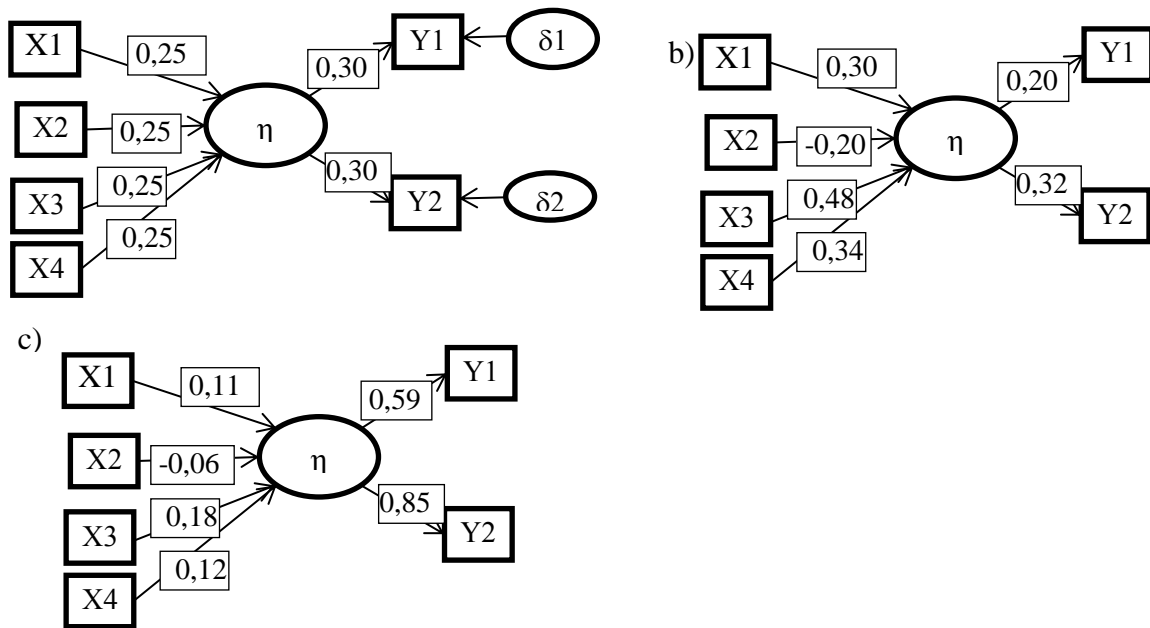
En conclusion, seuls des indicateurs dont on sait qu'ils sont susceptibles de n'intégrer qu'une faible erreur de mesure doivent être utilisés.

4.2.3.6 La gestion de la multi-colinéarité

Bien que théoriquement aucune contrainte ne soit imposée sur les covariances entre les mesures formatives, une forte colinéarité peut induire une instabilité des paramètres du modèle de mesure, particulièrement dans le cadre de CB-SEM et de l'approche PLS-SEM. C'est un problème usuel avec la régression multiple. Dans ce cas les paramètres estimés sont instables. La colinéarité peut être jugée excessive par le calcul d'un VIF généralement supérieur à 10. Il est préconisé dans ce cas de supprimer les indicateurs « redondants » tout en s'assurant que la signification postulée *à priori* du construit n'est pas altérée.

Nous illustrons ce problème par une simulation qui reprend le même cadre que la genèse du modèle de base posé Figure 21, mais ici nous avons imposé que les indicateurs formatifs soient fortement corrélés (corrélations de 0,9), avec toujours des poids théoriques des différents indicateurs identiques sur la latente. Le modèle théorique est Figure 25 a/, nous estimons les paramètres sous CB-SEM (Figure 25 b/) et sous PLS-SEM (Figure 25 c/). Nous voyons que la colinéarité entre les quatre indicateurs formatifs amène des résultats fortement biaisés : X2 semble influencer η de façon négative. Les indicateurs de fiabilités amèneraient pourtant le praticien à valider ce modèle. Il faut donc absolument vérifier la colinéarité potentielle des indicateurs formatifs.

Figure 25 : a) Modèle théorique avec colinéarité des X_i ; b) estimé avec CB-SEM ** ; c) estimé avec PLS-SEM **



* $\chi^2 : 3,20 ; p \text{ value} : 0,36 ; RMSEA : 0,011$

** $gof=0,251$

4.2.3.7 Le rôle médiateur du construit et les contraintes de proportionnalité

Une hypothèse encore plus contraignante est celle qui stipule que les indicateurs formatifs déterminent uniquement la variable latente formative. Le schéma d'identification (Figure 19), qui nécessite l'ajout de variables endogènes, conduit implicitement à imposer des contraintes sur les paramètres du modèle de mesure (Bollen & Davis, 2009a; Franke et al., 2008). Cette contrainte peut être formalisée comme suit dans le cadre de l'ASC (il faut enlever le terme d'erreur pour PLS-SEM) :

$$\eta = \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \zeta \quad (21)$$

$$\eta_I = \lambda_I \eta + \delta_I \quad I=1,2,\dots,Q \quad (22)$$

En substituant $\eta = \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \zeta$ (21) dans $\eta_i = \lambda_i \eta + \delta_i$ $i=1,2,\dots,q$ (22), nous obtenons

$$\eta_I = \lambda_I (\gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4) + \lambda_I \zeta + \delta_I \quad (23)$$

De l'équation $\eta_i = \lambda_i (\gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4) + \lambda_i \zeta + \delta_i$ (23), il ressort que l'effet de la VLC est égal à λ_i et celui d'un indicateur X_i est égal à $\lambda_i \gamma_i$. Tandis que l'effet relatif de X_i sur deux construits endogènes η_i et η_j est donné par

$$\lambda_i \gamma_i / \lambda_j \gamma_j \quad (24)$$

soit

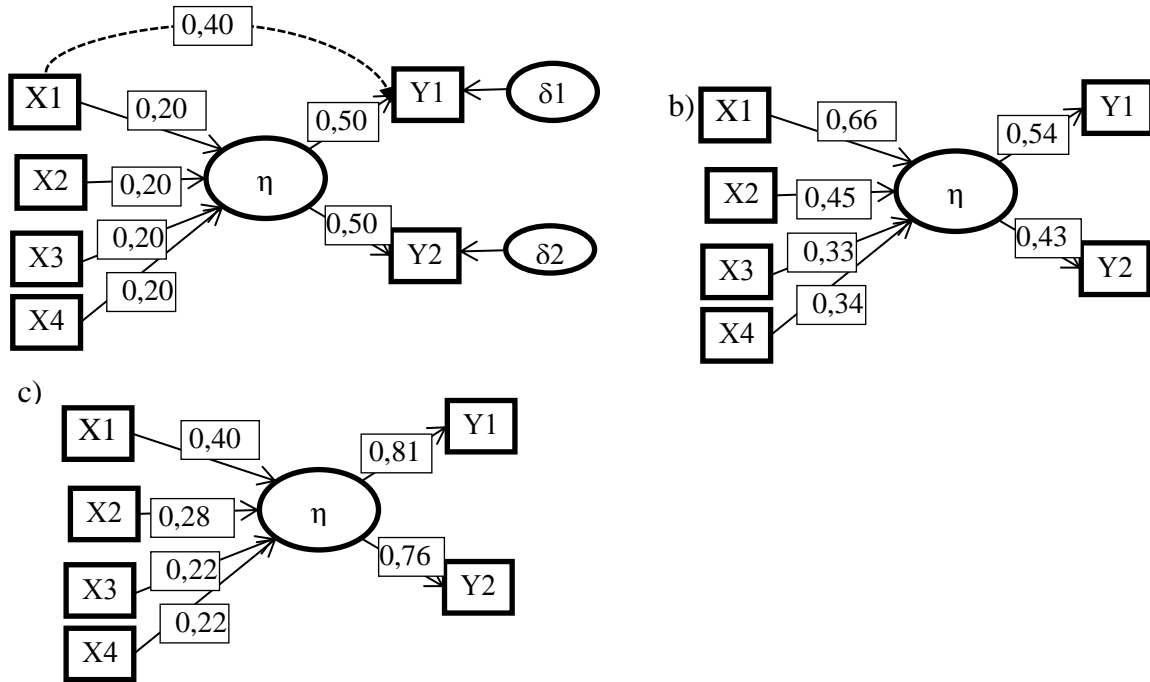
$$\lambda_i / \lambda_j \quad (25)$$

En conséquence les indicateurs formatifs X_i ont des effets proportionnels sur les construits endogènes η_i du même ordre de grandeur que celui de la VLC η qu'elles mesurent.

Formellement, il est facile de montrer que dans la Figure 19, lorsque la variable η influence par exemple 2 fois plus η_1 que η_2 , alors chaque mesure formative impacte aussi dans la même proportion η_1 et η_2 . Ainsi si la VLC a un impact 2 fois plus grand sur η_1 que sur η_2 , alors chacun de ses indicateurs X_i doit avoir un effet relatif deux fois plus grand sur η_1 que sur η_2 . Concrètement, cette propriété interdit que l'un des indicateurs formatif X_i du construit η puisse avoir un effet causal sur les variables endogènes Y_1 (ou η_1) et Y_2 (ou η_2) qui ne soit pas médiatisé par la variable latente η . Il convient donc de s'assurer que la variable latente est la seule variable médiatrice entre ses indicateurs et ses effets sur d'autres variables latentes endogènes. Une des solutions (proposée par Bollen & Davis, 2009a et Franke et al., 2008)) est d'utiliser les MI (« modifications d'indice » implanté dans tous les logiciels identiques à CB-SEM) qui révéleront si un lien « oublié » entre les indicateurs formatifs et d'autres variables (en dehors de la latente formative) sont significatifs (un seuil de 7.88 de l'amélioration du test d'ajustement χ^2 est proposé par Jöreskog & Sörbom (1993) au lieu d'un seuil usuel à 3.84 pour assurer une robustesse au test). Si tel est le cas, le modèle est alors mal spécifié... et il est crucial de ne pas ajouter ce lien en plus, mais il faut impérativement reconsidérer la définition de la latente formative. Le schéma ci-dessous (Figure 26) illustre par la flèche en pointillée une relation théoriquement interdite entre une mesure formative et une autre variable endogène. J'ai généré des données avec un tel modèle théorique et estimé ce modèle en faisant abstraction du lien entre X_1 et Y_1 . Les modèles estimés sous CB-SEM (Figure 26 b/) et PLS-SEM (Figure 26 c/) donnent des résultats biaisés : X_1 semble avoir un poids plus important qui est la conséquence du lien entre X_1 et Y_1 non introduit dans le modèle estimé. Le *gof* sous PLS-SEM apparaît très bon, le *RMSEA* sous CB-SEM est acceptable (<0.08). Seul CB-SEM permet de vérifier que les règles de proportionnalité sont respectées (par les tests sur les modifications d'indices). Dans notre

exemple, le logiciel informe que le lien entre X1 et Y1 apporterait une modification significative de la fiabilité du modèle (MI=9,41). Cela nous alerte sur la mauvaise spécification du modèle illustré par la Figure 26 b/, test impossible à faire avec PLS-SEM⁴¹ (Figure 26 c/).

FIGURE 26 : A) MODELE THEORIQUE AVEC UNE LIAISON FALLACIEUSE ; B) ESTIME AVEC CB-SEM ** ; C) ESTIME AVEC PLS-SEM **



* $Chi^2 : 11,97 ; p\text{ value} : 0,007 ; RMSEA : 0,077$

** $gof=0,46$

4.2.3.8 La gestion des covariances des indicateurs formatifs

Théoriquement, les indicateurs formatifs peuvent être corrélés entre eux. Avec CB-SEM, la gestion des covariances pose problème. Le nombre de covariances augmente exponentiellement en fonction du nombre de mesures formatives (Jarvis et al., 2003), or l'exhaustivité de la mesure implique le plus souvent un nombre important d'indicateurs (avec quatre indicateurs formatifs nous avons six covariances à estimer ; avec six indicateurs le nombre de covariances à estimer monte à quinze). De plus, les indicateurs formatifs peuvent covarier avec d'autres variables exogènes du modèle. Ceci rend l'estimation délicate dès que le modèle devient important (le nombre de paramètres à estimer devient trop important). Une

⁴¹ Nous rediscuterons des MI pour vérifier les contraintes de proportionnalités par simulations.

méthode avec CB-SEM pour contrôler les covariances est de suivre la pratique suivante (Jarvis et al., 2003; MacCallum & Browne, 1993) : estimer toutes les covariances entre tous les construits latents exogènes et les variables manifestes, puis estimer une série de modèles hiérarchiquement liés (nichés) qui permettent de libérer des covariances (en les spécifiant nulles). Pour tester si une covariance peut être retirée, il suffit sous CB-SEM de comparer les Chi2 entre les deux modèles (qui sont nichés) et de faire un test du Chi2 à un degré de liberté. Comme le praticien aura essayé de prendre des indicateurs non-corrélés, il pourra certainement vérifier que certaines covariances peuvent être retirées, ce qui réduira le nombre de paramètres à estimer et rendra l'ensemble des résultats plus fiables. Cette démarche n'est faisable que sous CB-SEM, mais PLS-SEM ne souffre pas de ce problème car l'algorithme n'estime pas le modèle d'un bloc. Par contre avec PLS-SEM, nous ne pouvons pas spécifier des corrélations nulles entre des indicateurs.

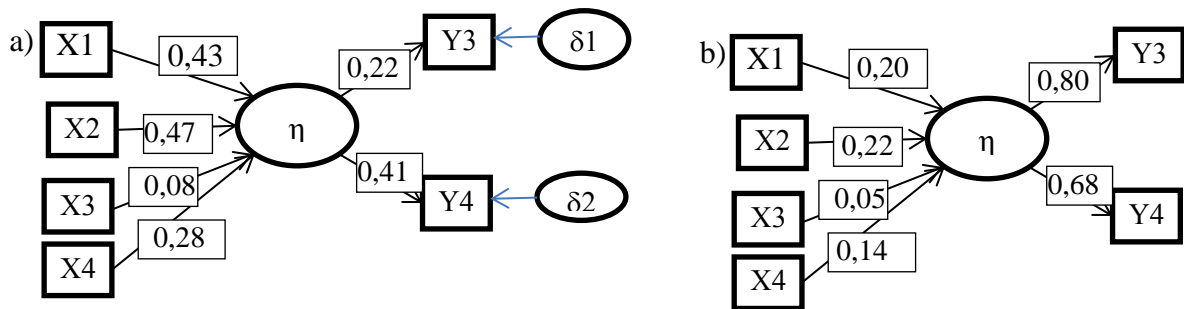
4.2.3.9 L'interprétation confondante

Ce problème est très clairement le plus important mis en avant par les opposants aux modèles formatifs. De nombreux travaux empiriques montrent que le choix des construits endogènes η_1 et η_2 , nécessaires à l'identification du SEM, peut altérer la signification du construit η . Ce phénomène a été mis en évidence dans la littérature à la fois dans le cadre de CB-SEM aussi bien que dans l'approche PLS-SEM (Howell et al., 2007b; G. Kim et al., 2010; Tensaout, Guyon, & Hanafi, 2011; Wilcox et al., 2008). Il est désigné sous le terme d'« interprétation confondante ». Nous souhaitons faire une présentation de ce phénomène qui est à l'origine de vifs débats entre les tenants et les opposants au modèle formatif.

Dans un SEM, l'interprétation confondante désigne la situation dans laquelle l'ajout ou la suppression d'une relation structurelle entre le construit η et d'autres construits η_i du SEM modifie significativement les valeurs des paramètres du modèle de mesure (formatif ou réflectif) du construit η , et par là altère sa signification postulée *à priori*, avant son estimation par le chercheur. Nous reviendrons plus longuement sur ce problème lors de nos simulations. Nous illustrons juste ce problème en reprenant le modèle de base (Figure 21). Nous modifions ce modèle en mettant deux nouvelles variables réflectives endogènes Y3 et Y4 à la place de Y1 et Y2 (Figure 27). η a théoriquement la même signification dans les deux modèles car η est estimé formativement par les mêmes X_i exogènes. Or nous voyons que les indicateurs formatifs ont des coefficients différents dans les deux modèles, et ceci avec CB-SEM et PLS-SEM. Notamment, la variable X3 qui était la plus influente dans le modèle initial (Figure 21)

devient ici non significative (Figure 27) et un praticien la retirerait de son modèle. Comment dès lors considérer avoir le même construit η dans les deux modèles (Figure 21 ou Figure 27) ?

FIGURE 27 : MODELE DE BASE AVEC Y3 ET Y4 A) ESTIME AVEC CB-SEM * ; B) ESTIME AVEC PLS-SEM **



* $Chi^2 : 3,53 ; p \text{ value} : 0,3 ; RMSEA : 0,018$

** $gof = 0,252$

Je rediscuterai avec des simulations du cadre empirique qui permette à un praticien de considérer que son SEM ne souffre pas d'interprétation confondante.

4.2.3.10 Le cas d'un construit endogène

La littérature académique discute essentiellement des construits exogènes avec des mesures formatives. En réalité, ce construit peut être aussi endogène, c'est-à-dire lui-même impacté par d'autres construits réflectifs ξ exogènes qui ne sont pas ses antécédents (Figure 28). Cette situation complique davantage les choses car cette fois la variable latente formative est déterminée à la fois par ses mesures et par d'autres construits exogènes. Cette spécification pose d'importantes questions conceptuelles. Pour rappel, les indicateurs formatifs représentent l'ensemble des causes du construit. Comment alors ξ peut influencer η sachant que cette dernière est essentiellement déterminée par ses indicateurs ? Si tel est le cas, cela reviendrait à considérer que les indicateurs formatifs ne représentaient pas l'exhaustivité des causes de η ... et donc que le modèle de mesure est mal spécifié. Cadogan & Lee (2013) et Tensaout & Guyon (2013) proposent une solution pour éviter ce « paradoxe » qui consiste à spécifier des relations causales entre les antécédents ξ et les indicateurs de mesures X_i de η (Figure 31). Mais comme l'explique alors Rigdon (2014), on ne peut considérer que ξ influence η car cela reviendrait à considérer que l'erreur de la latente, qui représente entre

autres l'influence de ξ sur η , serait liée aux indicateurs de η (cf. le point précédent sur l'exogénéité de l'erreur). Nous nous trouvons devant un autre paradoxe : une latente formative ne pourrait pas être intégrée comme une variable endogène avec des variables latentes exogènes dans un SEM. La discussion à ce sujet est à peine initiée dans la littérature académique, mais la prudence préconiserait de ne pas utiliser de latentes formatives endogènes dans des SEM. Ce point renvoie donc à une lecture critique de bien des travaux où les praticiens utilisent sans précautions des variables latentes formatives endogènes.

FIGURE 28 : CONSTRUIT ENDOGENE AVEC DES INDICATEURS FORMATIFS QUI POSE LE PARADOXE DE LA MAUVAISE SPECIFICATION DU MODELE DE MESURE.

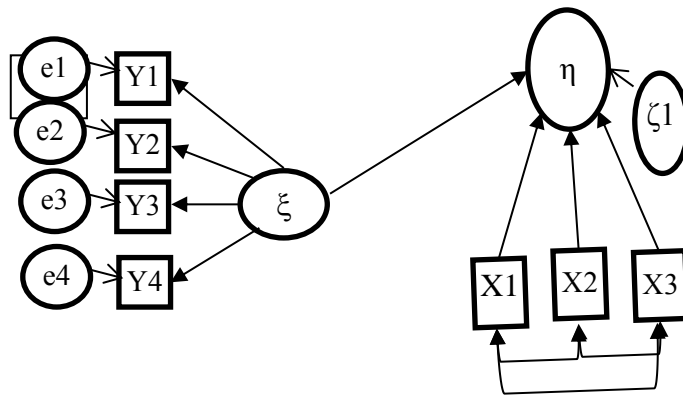
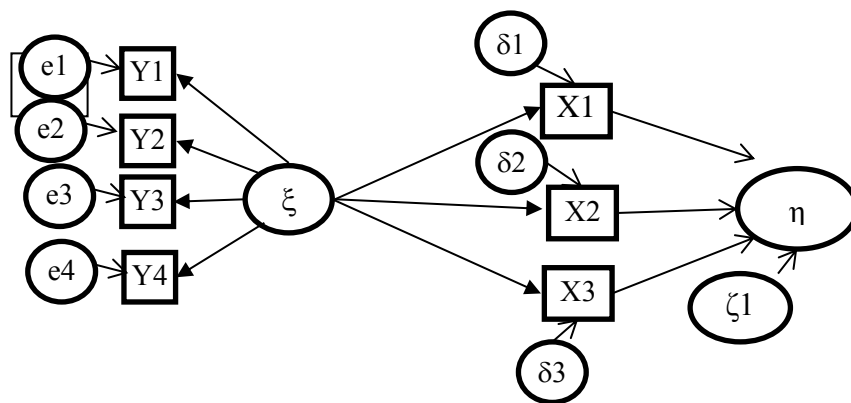


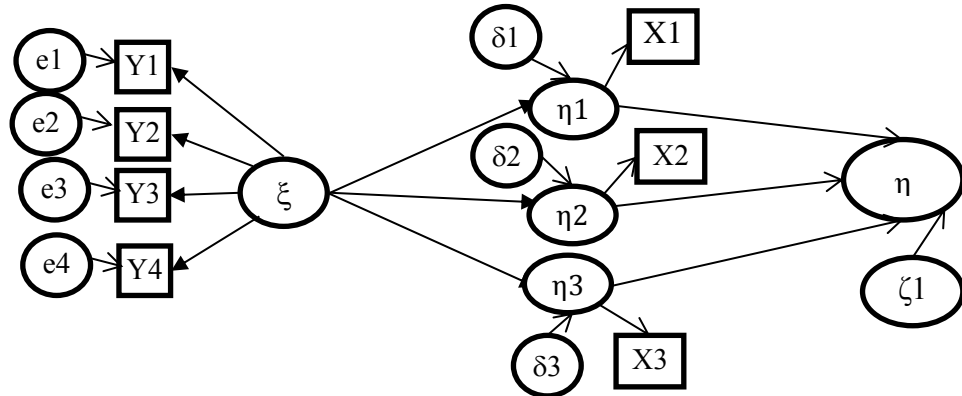
FIGURE 31 : CONSTRUIT AVEC DES INDICATEURS FORMATIFS ENDOGENES QUI POSE LE PARADOXE DE LA VARIABLE LATENTE FORMATIVE NON DIRECTEMENT INFLUENCEE PAR D'AUTRES VARIABLES LATENTES.



Temme, Diamantopoulos, & Pfgfeidel (014) ont proposé récemment (Figure 30) de poser le construit formatif endogène comme un construit de second ordre relié formativement à des

construits de premier ordre (ses indicateurs). Cette formalisation semble résoudre tous les problèmes méthodologiques... mais est trop récente pour valider son opérationnalité réelle.

FIGURE 30 : MODELISATION D'UNE VARIABLE LATENTE FORMATIVE



Tiée de Temme et al. (2014)

4.2.3.11 Les construits formatifs de second-ordre

Les construits de second ordre sont des construits “plus abstraits” (Chin, 1998; Jarvis et al., 2003). Je renvoie au papier de Jarvis et al. (2003) pour les différentes configurations d’un construit de second ordre. Comme MacKenzie, Podsakoff, & Jarvis (2005) l’expliquent, le processus de validation d’un construit de second ordre est presque identique au processus d’un construit de 1^{er} ordre. Cependant, un construit de second ordre pose le problème des erreurs associées aux construits (Diamantopoulos et al., 2008) et nous devons le plus souvent utiliser des “variables fantômes » (Rindskopf, 1984), ce qui est problématique (MacCallum & Browne, 1993). En conséquence je ne développerai pas de telles configurations.

4.2.4 Conclusion

J’ai souligné que les débats sur la légitimité de l’usage d’un modèle de mesure formatif d’un construit ont fait l’objet de nombreuses publications. J’ai montré qu’effectivement le développement des mesures formatives nécessite un ensemble de conditions très difficiles à satisfaire : absence d’erreur de mesure, exhaustivité des mesures, indépendance locale des variables exogènes. A ces difficultés s’ajoutent celles communes à tout travail de modélisation comme l’équivalence observationnelle, l’omission de variables importantes comme les variables confondantes etc.

Les opposants aux modèles formatifs pointent le fait que les bases théoriques et conceptuelles du modèle de mesure formatif sont fragiles, notamment l'absence de l'erreur de mesure, l'hypothèse d'exhaustivité des mesures, et les contraintes de spécification des mesures nécessaires à l'identification. Mes exemples simulés montrent à l'évidence que le développement d'un modèle formatif valide est une tâche difficile et peut donc conduire à des résultats fallacieux. PLS-SEM n'est pas moins sensible à l'omission de mesures importantes, à la colinéarité et à l'interprétation confondante que CB-SEM.

Aussi, pour les chercheurs ne pouvant éviter d'utiliser un modèle de mesure formatif sans vouloir avoir recours à ces méthodes, je discute par simulations Monte Carlo de la robustesse des estimations de ce modèle sur un spectre large de paramètres à même de pouvoir donner des indications sur le cadre paramétrique où un tel modèle formatif peut être considéré méthodologiquement fiable (sans poser sa validité ontologique).

4.3 Simulations

Je développe par la suite des simulations Monte Carlo pour analyser le cadre paramétrique dans lequel un modèle de mesure formatif peut paraître efficient. Je me situe donc toujours dans le cadre du réalisme-empirique et considère comme acceptable de poser une propriété mentale comme potentiellement estimable par un modèle formatif.

4.3.1 Le cadre des simulations

Toute la discussion préliminaire renvoie bien des questions sur l'opérationnalisation de modèles de mesure formatifs même dans un cadre empiriste-réaliste. Cependant, comme je vais le discuter, la littérature a effectué in fine peu de simulations de tels modèles. Ce travail développé ici cherche à apporter de nouvelles réponses pour clarifier le débat sur les construits formatifs. L'objectif est de répondre à un certain nombre de questions à partir de simulations Monte Carlo. Cette méthode est pertinente lorsque le praticien ne dispose pas des priorités statistiques sur les distributions des paramètres ((Boomsma, 2013; Chin, Thatcher, & Wright, 2012; Goodhue et al., 2012; Marcoulides & Chin, 2013; Paxton, Curran, Bollen, Kirby, & Chen, 2001).

Les simulations Monte Carlo partent d'un modèle théorique, puis en imposant des contraintes (modifications des valeurs des paramètres du SEM que j'appellerai : cadre paramétrique) nous spécifions différentes configurations du SEM. L'objectif est de vérifier la robustesse des estimations sur un nombre de simulations suffisant par cadre paramétrique (la littérature retient en général 500 simulations par cadre paramétrique). Je suivrai les recommandations de (Marcoulides & Chin, 2013; Paxton et al., 2001) pour le design de les simulations : simulations du vrai modèle puis estimations des paramètres.

J'informe que je ne donne que des résultats synthétiques pour rendre ce travail lisible. La reproduction de l'exhaustivité des résultats aurait été impossible en format papier, de plus un tel report aurait noyé le lecteur dans un nombre de tableaux trop important.

4.3.2 Le modèle théorique

Des exemples ne peuvent refléter l'ensemble des situations. De plus, la robustesse de certains modèles relativement à de mauvaises spécifications ne peut être testée que par simulations. Il est donc primordial de poser des simulations comparant un résultat théorique (le modèle

simulé) et un résultat empirique (le résultat estimé) sur un spectre large de paramètres et avec suffisamment de simulations pour avoir une certaine robustesse des résultats (500 répliques est le cadre usuel).

Il faut donc partir d'un cadre théorique et non d'exemples comme G. Kim et al. (2010) car nous n'avons aucune idée du « vrai modèle » au-delà du cadre conceptuel ; et considérer un modèle « théoriquement » valide ne veut pas dire qu'au niveau des données il le soit. De plus, même des simulations doivent l'être dans un cadre hypothétique, c'est-à-dire en ayant posé un cadre de simulations correspondant au cadre théorique. Pour exemple Howell et al. (2007) et Wilcox et al. (2008) simulent des données mais ne donnent que les matrices de covariances sans spécifier le modèle théorique qui a permis de générer de telles données. Comme je le montrerai ces exemples proposés par ces auteurs sont critiquables.

Il faut donc avoir un cadre « *theory driven* » (Bainter & Bollen, 2015). C'est-à-dire construire des simulations correspondant au modèle que nous voulons vérifier. Si les exemples de G. Kim et al. (2010) ou Franke et al. (2008) peuvent être considérés comme basés sur des données renvoyant à un cadre théorique, c'est un cadre théorique conceptuel et non pas un cadre théorique de données simulées. Or il faut connaître la structure des données générées (et non pas le cadre conceptuel de leur recueil) pour vérifier par simulations la robustesse de modèles statistiques.

Je me propose donc de simuler dans une approche « *theory driven* ».

4.3.3 Les lacunes des simulations dans la littérature

Le développement d'un modèle de mesure formatif valide nécessite le respect de plusieurs hypothèses contraignantes : exhaustivité des mesures et respect de la règle de proportionnalité. La violation de l'une de ces hypothèses représente une menace sérieuse à la validité.

4.3.3.1 La règle de proportionnalité

Formellement la règle de proportionnalité se traduit par des contraintes sur les paramètres du SEM (Bollen & Davis, 2009a; Franke et al., 2008) : les indicateurs formatifs exogènes ont la même influence (proportionnellement) sur les mesures réflectives de la variable latente (Figure 18) ou sur les variables latentes endogènes (Figure 19). En conséquence la violation de ces contraintes est une indication importante d'une mauvaise spécification du modèle de mesure. Bollen & Davis (2009a) proposent entre autres de tester ces contraintes de

proportionnalité par le calcul des MI (« modifications d’indices ») en évaluant l’augmentation de la valeur du Chi 2 lorsque on introduit des liens directs entre les indicateurs formatifs et les indicateurs (ou latentes) réflectifs (voir plus loin).

Le respect de ces contraintes de proportionnalité sont ignorées par Howell et al. (2007b); G. Kim et al. (2010); Wilcox et al. (2008) et explique, à notre avis, au moins partiellement l’instabilité observée des paramètres estimés par ces auteurs. Comme le montre le tableau 2, les MI sont largement supérieurs au seuil admis pour ces exemples ($MI < 7.88^{42}$, Jöreskog & Sörbom, 1993)), ce qui montre que l’ajout de liens directs entre les indicateurs formatifs et les autres variables latentes endogènes peut améliorer sensiblement le fit du SEM. Par conséquent dans tous ces exemples la variable latente η ne médiatise pas entièrement les effets des indicateurs formatifs sur les autres variables endogènes du SEM. Ce résultat montre à l’évidence que la règle de proportionnalité n’est pas satisfaite ce qui caractérise des modèles mal spécifiés, et explique en partie l’instabilité des estimations obtenues par ces auteurs.

TABLEAU 5 : LES MI DES EXEMPLES DE (HOWELL ET AL., 2007A) ET G. KIM ET AL. (2010)

	(Howell et al., 2007b)			Kim et al. (2010)		Wilcox et al. 2008(**)		
	Ex. 1 (Table 2 p. 211)	Ex. 2 (Table 3, p. 212)	Ex. 3 (Table 4, p. 213)(*)	Case 1, F- model 1 (p. 351)	Case 1, F- model 2 (p. 351)	Ex. 1 (fig 3, p. 1223)	Ex. 2 (fig 4, p. 1224)	Ex. 1 (fig 5, p. 1225)
Nombre de MI > 7.88 entre des indicateurs formatifs et des variables manifeste et/ou latentes réflectives	8	6	6	1	1	4	0	4

(*) Les estimations que j’ai trouvées diffèrent de celles reportées par Howell et al. (2007a) (j’ai fait les calculs avec R et MPlus).

(**) La taille de l’échantillon n’est pas précisée, j’ai utilisé un échantillon de 500 comme Bainter & Bollen (2014).

⁴² Je pose le seuil à 7.88 comme le proposent (Jöreskog & Sörbom, 1993) et non 3.84. Cela permet de caractériser des modèles réellement déviants.

4.3.3.2 Exhaustivité des indicateurs formatifs

L'autre source qui menace la validité des mesures formatives est la non-exhaustivité des indicateurs de mesure. Formellement, celle-ci implique théoriquement que la variance de l'erreur associée à la latente formative soit faible, qui se traduit par un SMC (*squared multiple correlations*) proche de 1 (hypothèse nulle). Cette erreur associée à la variable latente formative est ignorée dans les exemples proposés par Franke et al. (2008); G. Kim et al. (2010). G. Kim et al. (2010) ne donnent pas les SMC, je les ai calculés et obtenu des valeurs considérées comme en dessous d'un seuil acceptable (modèle F-model 1 et F-model 2 de leurs articles, estimations basées sur la matrice de corrélation donnée par ces auteurs dans leur annexe B). Ces résultats montrent que les erreurs associées aux latentes formatives sont élevées, ce qui montre que les indicateurs formatifs ne sont pas exhaustifs, et donc que les modèles sont mal spécifiés.

TABEAU 6 : SMC DES EXEMPLES DE LA LITTÉRATURE

	Franke et al. (2008)		(G. Kim et al., 2010)	
	Ex. 1	Ex. 2	Case 1, F-model 1 (p. 351)	Case 1, F-model 2 (p. 351)
SMC	.02	.11	.34	.35

Notons que les SMC des exemples développés par Howell et al. (2007b) sont à des seuils acceptables (respectivement 0.64, 0.79 et 0.89 pour les exemples 1, 2 et 3).

4.3.3.3 Les coefficients standardisés

Comme je l'ai expliqué précédemment, la valeur des coefficients dépend du choix de normalisation utilisé (Diamantopoulos, 2011; Franke et al., 2008). De ce fait, seuls des coefficients standardisés permettent de pouvoir analyser les résultats sans risque d'erreur d'interprétation. Bainter & Bollen (2014) proposent des simulations avec des liens structuraux (les liens entre la latente formative et les latentes réflectives) annoncés comme « medium » (coefficient de 0.5) et « high » (coefficient de 0.8). A partir du code MPlus de ces simulations (que Bainter m'a envoyé), j'ai repris leurs exemples en standardisant les résultats. Ce que Bainter et Bollen ont considéré comme « medium » est en réalité le coefficient non-standardisé entré dans MPlus (0.5) mais sans tenir compte de l'effet de normalisation. En standardisant, ce coefficient devient 0.78. C'est-à-dire que le coefficient standardisé

« medium » est au niveau de ce que ces auteurs considèrent être un coefficient élevé. De plus le SMC est supérieur à 0.80, ce qui représente un SMC élevé.

En conclusion, le cadre paramétrique utilisé par Bainter et Bollen correspond à un cadre très restrictif de modèles : liens structuraux importants et SMC élevé. Kuncel, Ones, & Hezlett (2001) par exemple ont fait une méta-analyse des corrélations dans les modèles structuraux et ces corrélations sont en général entre 0.1 et 0.5. Le cadre paramétrique de Bainter et Bollen est donc largement en dehors du cadre pratique usuel.

4.3.3.4 PLS-SEM vs CB-SEM

Le problème de l'approche (PLS-SEM vs CB-SEM) pour estimer un modèle formatif est discuté dans la littérature (Diamantopoulos, 2011; Gefen et al., 2011; Goodhue et al., 2012, 2012; Marcoulides, Chin, & Saunders, 2009; Qureshi & Compeau, 2009; Ringle et al., 2012). Différents auteurs considèrent que PLS-SEM est préférable à CB-SEM (Gefen et al., 2011). Mais G. Kim et al. (2010) ont montré par exemple que PLS-SEM ne semblait pas contourner les biais potentiels de CB-SEM. De plus PLS-SEM ne donne pas d'indicateur *goodness-of-fit* alors que (Bollen, 2011) discute de la nécessité d'utiliser de tels indicateurs pour valider un modèle formatif. De ce fait il est important de voir la sensibilité de PLS-SEM relativement à CB-SEM. Ringle et al. (2009) et Goodhue et al. (2012) qui ont modélisé des simulations avec des modèles formatifs et PLS-SEM, mais sans examiner la stabilité des estimations.

Ce travail intègre donc des simulations avec PLS-SEM. Mais la discussion sur PLS-SEM couvre actuellement de vastes débats et il faudrait un travail spécifique pour poser tous les problèmes et résultats liés à cette approche. Comme PLS-SEM est, de par la nature du modèle, mathématique posé en contradiction avec le cadre empiriste-réaliste (il n'y a pas d'erreur de mesure de la latente, on a donc un index et non une variable latente au sens de CB-SEM), je ne m'étendrai pas sur PLS-SEM ; pourtant cette discussion mériterait d'être posée car des articles critiquent ce regard sur PLS-SEM (Rigdon, 2012)... mais un tel travail nécessiterait une nouvelle thèse.

4.3.3.5 Conclusion

Aucune des simulations ou exemples donnés par la littérature sur les modèles formatifs n'est valide pour pouvoir généraliser les résultats. Il y a une nécessité de construire de véritables simulations Monte Carlo. C'est la suite de ce travail.

4.3.4 Design des simulations Monte Carlo

Je donne ici le design des simulations. L'objectif est de construire des modèles théoriques qui correspondent au cadre hypothétique d'une latente formative, puis de vérifier la robustesse des estimations sur un spectre large de cadres paramétriques. Ces simulations permettront de vérifier dans quel cadre un modèle formatif peut être considéré comme fiable en termes d'estimations.

Dans un second temps, je générerai des modèles « mal spécifiés » : 1/ ne respectant pas les contraintes de proportionnalités ; 2/ ne respectant pas l'exhaustivité des indicateurs. L'objectif sera de voir si les indicateurs statistiques permettent de détecter ces problèmes.

4.3.4.1 Les modèles simulés

Les SEM dans lesquels sont discutés les modèles de mesure formatifs usuellement dans les applications empiriques sont : a/ un modèle MIMIC avec 4 indicateurs formatifs + 2 indicateurs réfléchifs, Figure 31 (Bagozzi, 2011b; Diamantopoulos, 2011) ; b/ MIMIC avec 4 indicateurs formatifs et 3 variables réfléchifs (Cadogan, Souchon, & Procter, 2008; Hardin, Chang, Fuller, & Torkzadeh, 2011) ; c/ un SEM comportant un modèle de mesure avec 4 indicateurs formatifs et 2 autres modèles de mesure réfléchifs de deux variables latentes endogènes η_1 et η_2 (avec chacun 4 manifestes), Figure 32 (Bainter & Bollen, 2014; Howell et al., 2007b; G. Kim et al., 2010; Wilcox et al., 2008) ; d/ un SEM comportant un modèle de mesure avec 4 indicateurs formatifs et 4 autres modèles de mesure réfléchifs de quatre variables latentes endogènes η_1 , η_2 , η_3 et η_4 (avec chacun 4 manifestes), Figure 33 (Bainter & Bollen, 2014; Howell, 2014).

FIGURE 31 : MODELE MIMIC

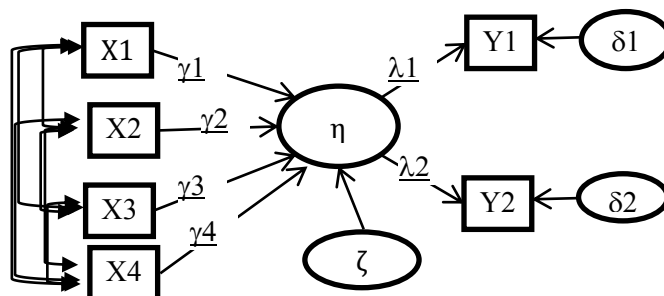


FIGURE 32: MODELE 1, VARIABLE LATENTE FORMATIVE η RELIEE A DEUX LATENTES REFLECTIVES η_1 ET η_2

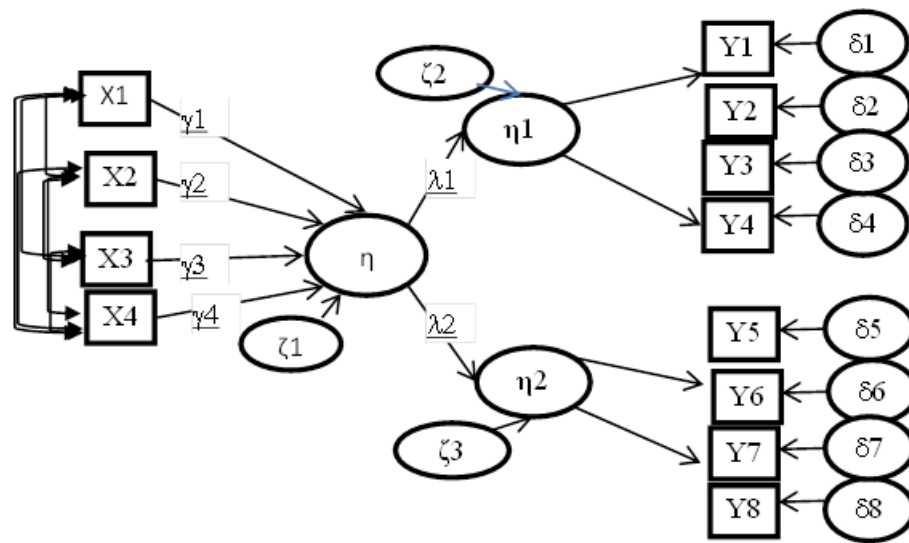
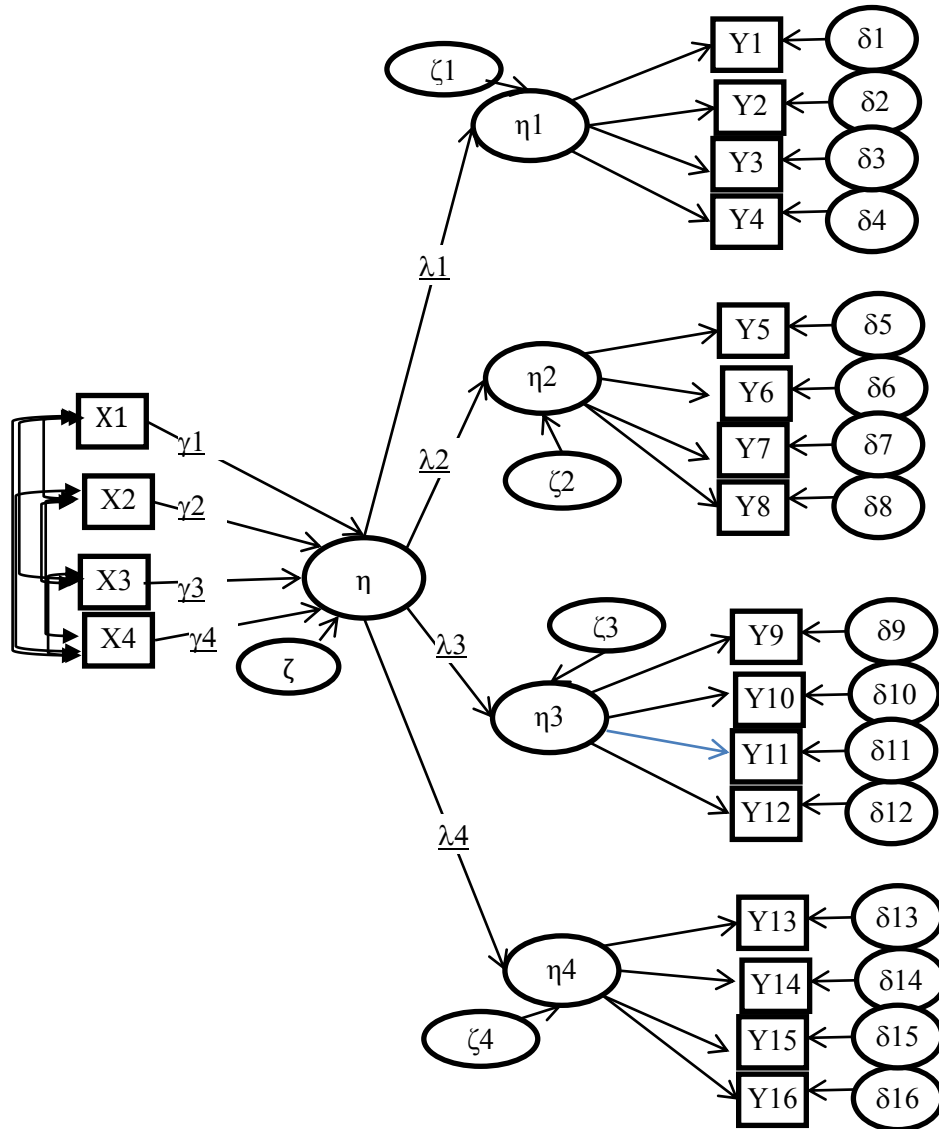


FIGURE 33 : MODELE 2, VARIABLE LATENTE FORMATIVE η RELIEE A QUATRE LATENTES REFLECTIVES η_1 A η_4



Nous proposons de discuter dans cet article à partir : 1/ MODELE 1 : SEM avec 4 indicateurs formatifs et 2 construits réflectifs ayant chacun 4 manifestes (Figure 32), qui correspond au modèle discuté par Howell et al. (2007b) ; 2/ MODELE 2 : SEM avec 4 indicateurs formatifs et 4 construits réflectifs ayant chacun 4 manifestes (Figure 33), qui correspond au modèle discuté par Bainter & Bollen (2014). En effet, un modèle MIMIC (Figure 31) sert plus de cadre de discussion académique que de cadre pratique (je renvoie à ce que j’ai développé sur les modèles MIMIC en introduction). Les praticiens intégrant des modèles de mesure formatifs dans leurs modèles risquent bien plus de devoir se rapporter aux résultats du modèle de la Figure 32 que du modèle de la Figure 31. Par ailleurs les résultats obtenus avec un modèle MIMIC (Figure 31) corroborent ceux obtenus avec le modèle SEM (Figure 32).

Cependant, pour vérifier certaines hypothèses, nous simulerons des modèles MIMIC pour simplifier le travail d'analyse.

4.3.4.2 Design pour les modèles biens spécifiés

Les simulations ont été faites en deux phases distinctes. Les jeux de données sont générés à partir d'un modèle théorique de référence : modèle 1 (Figure 32) ou Modèle 2 (Figure 33). Pour la phase 1 des simulations, seul le modèle 1 a été utilisé ; pour la phase 2 les modèles 1 et 2 ont été retenus.

Pour les phases 1 et 2, en se basant sur les applications empiriques publiées, nous avons retenu plusieurs configurations en fonction notamment de la taille de l'échantillon, noté N par la suite ; du niveau de la variance expliquée par les indicateurs formatifs de la latente η , noté SMC par la suite ; des liens standardisés entre cette latente η et les autres variables endogènes réflexives (manifestes ou latentes), noté λ par la suite.

Pour la phase 1, la colinéarité entre les indicateurs formatifs, noté R par la suite, avait été utilisé. Pour la phase 2, afin d'alléger le plan nous n'avons retenu que des indicateurs non corrélés ($R=0$).

Dans toutes nos simulations, nous posons :

- dans chacune des configurations, les paramètres γ des indicateurs formatifs ont le même poids sur la latente ($\gamma_1=\gamma_2=\gamma_3=\gamma_4$). Mais le niveau des derniers peut différer d'une configuration à l'autre.
- Les effets directs λ_i de la latente η sur les autres variables endogènes sont posés égaux. Mais leur niveau varie d'une configuration à l'autre.
- les distributions des variables du SEM sont gaussiennes centrées réduites.
- Les corrélations entre les indicateurs réflexifs Y_i des variables latentes endogènes sont fixées à 0,9.

Les paramètres qui varient dans nos simulations sont :

- N (tailles de l'échantillon). 1/ Phase 1 : 200, 300, 500, 750, 1000 ; 2/ Phase 2 : 250, 500, 1000. Des simulations sur des échantillons de 100 ont été testées et apportent un nombre trop important de modèle CB-SEM qui ne convergent pas pour pouvoir être analysés.
- SMC (variance de la latente η expliquée par les indicateurs formatifs) : la littérature sur les SEM recommande des SMC supérieurs à 0,5. 1/ Phase 1 : Nous fixons un SMC variant entre 0,6 à 0,9 avec un pas de 0,1 (soit 0,6 ; 0,7 ; 0,8 ; 0,9). 2/ Phase 2 : Nous fixons un SMC variant entre 0,5 à 0,9 avec un pas de 0,1.

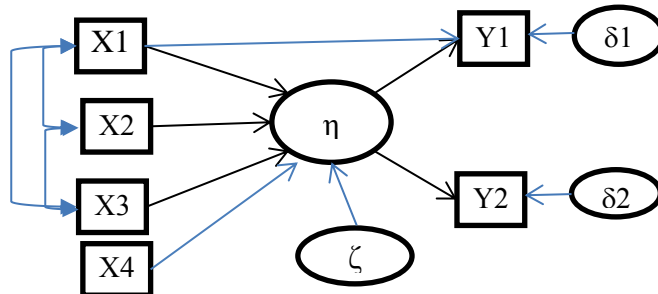
- λ (les liens standardisés entre la latente formative η et les variables réfléchives ou endogènes η_1 et η_2), varie (phase 1 et 2) de 0,2 à 0,8 (avec un pas de 0.2) pour toutes nos hypothèses ;
- R (colinéarité entre les indicateurs formatifs). 1/ Phase 1 : varie de 0 à 0.8 par pas de 0.2 ; Phase 2 : posé à 0.

4.3.4.3 Design pour les modèles mal spécifiés

Je discuterai des moyens empiriques pour détecter une mauvaise spécification du modèle de mesure formatif. Pour des raisons de simplicité et de compréhension, nous présentons les résultats du MIMIC (Figure 31). Comme nos hypothèses sont basées sur des indicateurs statistiques que seul CB-SEM fournit, nous ne ferons des simulations que dans ce cadre.

Pour vérifier l’hypothèse de proportionnalité, nous vérifions par simulations si la violation des règles de proportionnalité est empiriquement détectable par les MI. Nous spécifions incorrectement un modèle MIMIC (Figure 31) en ajoutant un lien entre X1 et Y1 (Figure 34):

FIGURE 34 : MODELE MIMIC AVEC LE NON-RESPECT DES REGLES DE PROPORTIONNALITE.



Le lien entre X1 et Y1 est donnée par l'équation :

$$Y_1 = (1 - \alpha) \times \eta + \alpha * X_1 + \varepsilon \quad (26)$$

Avec α allant de 0.1 à 0.8 par pas de 0.1.

Le pourcentage de lien total de X1 avec Y1 qui n'est pas médiatisé par la latente η varie de 10% à 80%. Pour nos simulations, nous posons R=0 (corrélations nulles entre les indicateurs formatifs) afin de ne pas perturber les résultats par des covariances entre ces indicateurs formatifs et une taille d'échantillon N=500. Nous prenons un SMC de de 0.6 à 0.9. Les paramètres λ_1 et λ_2 dans la Figure 18 représentent les corrélations entre η et, respectivement, Y1 et Y2. Dans le modèle théorique (Figure 18), c'est la corrélation globale (englobant

conjointement η et X_1) avec Y_1 qui varie de 0.2 à 0.8 ; λ_2 varie de 0.2 à 0.8. Nous rejetons l'hypothèse nulle (respect des contraintes de proportionnalité), lorsque les Modifications d'indices fournis par CB-SEM sont supérieures à 4⁴³.

Pour examiner l'exhaustivité des indicateurs, nous retenons uniquement le modèle MIMIC (Figure 18) avec les mêmes paramètres que pour les modèles biens spécifiés, mais avec un SMC de la variable latente η variant de 0.1 à 0.5.

4.3.5 Critères

L'analyse de la fiabilité d'une mesure formative ne peut être testée que sur la base de critères, or ceux utilisés par la littérature académique posent problème et c'est une des grosses lacunes de tous les travaux académiques jusqu'à présent.

4.3.5.1 Introduction

La littérature n'a considéré comme seul critère jusqu'à présent que l'instabilité des coefficients des indicateurs formatifs (les γ_i) mais sans user de tests ou intervalles de confiance pour vérifier ces variations en termes de significativité. Il apparaît crucial d'user de méthodes statistiques validant si ces instabilités sont significatives et donc problématiques.

De plus, la colinéarité entre les indicateurs formatifs peut très bien faire varier les coefficients estimés (identiquement à la régression multiple, Diamantopoulos & Sigauw, 2006), et donc on ne peut conclure sur l'efficacité des estimations d'un modèle de mesure en nous basant uniquement sur des estimations de ces indicateurs formatifs. C'est le problème essentiel que nous (Mouloud Tensaout et moi-même, Guyon & Tensaout, 2015) avons soulevé sur les critères utilisés dans la littérature académique. L'objectif d'un modèle de mesure est de mesurer correctement une variable latente. Les coefficients formatifs ne peuvent être analysés (à cause de la colinéarité), et il faut en réalité analyser si la variable latente estimée est « proche » de la variable latente théorique.

Enfin, le cœur du problème d'un modèle est de valider si la variable latente formative a une influence estimée sur les variables latentes réflectives correspondant au cadre théorique.

⁴³ Il mesure la baisse du Chi-deux, lorsque un des paramètres du modèle n'est pas fixé à 0 (p.e. ajouter une relation dans le MES).

4.3.5.2 Critère sur les coefficients formatifs

La discussion académique est posée uniquement sur la base de « modification » des estimations des paramètres γ du modèle de mesure. Il nous paraît nécessaire de poser à priori une règle de décision plus précise qui tient aussi en compte des écart-types des estimations des paramètres γ et non seulement de leurs niveaux. Cette procédure est standard dans la conduite des tests statistiques et permet d'apprécier si le sens empirique diverge du sens nominal au niveau des indicateurs formatifs. Comme nous travaillons sur des simulations, le sens nominal est connu : c'est le modèle théorique posé. Pour rappel, dans nos modèles théoriques (Figure 32 et Figure 33) les indicateurs formatifs ont le même poids γ sur la latente η . Nous considérons que le sens empirique diverge du sens nominal lorsque l'une des conditions ci-dessous n'est pas satisfaite :

- ✓ Les indicateurs ont des coefficients significatifs (Bainter & Bollen (2015, p.69) : « *If an indicator has face validity as a measure of the latent variable, but the empirical estimates suggest otherwise (i.e, weak, zero, or opposite sign), this would cast doubt on the validity of the measure* »)
- ✓ Les coefficients estimés ne divergent pas des coefficients théoriques. Comme les γ_i sont pris identiques et que c'est le poids relatif de ces coefficients qui nous intéresse, cette condition revient à vérifier que tous les coefficients sont identiques... ou du moins non-différents de façon significative. Avec CB-SEM, la valeur brute de l'estimation ne nous intéresse pas car elle dépend de la normalisation prise ; de même avec PLS-SEM il y a une normalisation des données et seule la comparaison des coefficients nous intéresse.

Ces deux critères permettent de valider que le modèle estimé est en phase avec le modèle théorique. Cependant, comme expliqué antérieurement, avec CB-SEM les tests de Student ou intervalles de confiance sur les indicateurs formatifs souffrent de biais (algorithmiques) qui rendent impossible de les utiliser comme test (Diamantopoulos, 2011; Franke et al., 2008). Franke et al. (2008) proposent d'utiliser des tests basés sur les *Likelihood Ratio* (LR), ce qui revient à faire des tests du χ^2 entre deux modèles (un contraint vs un libre). Concrètement les coefficients des indicateurs formatifs sont testés par LR pour savoir s'ils sont tous différents de zéro (un test par indicateur) ; et un test où le modèle contraint pose les indicateurs comme ayant des coefficients identiques vs un modèle libre sur ces coefficients. Ces tests sont contrôlés par Bonferonni ($\alpha = 0.05/5$).

Sous PLS-SEM, du fait que cette méthode ne fournit pas d'erreur associée à la latente η mesurée formativement, seuls des intervalles de confiance empirique (par Bootstrap par exemple) sont utilisables pour les paramètres γ . Nous avons donc effectué ces tests à partir de Bootstrap (100 répliques) avec PLS-SEM.

Nous utilisons une statistique de comptage : pour chaque configuration, nous retenons les simulations qui satisfont les conditions énumérés ci-dessus. Si 95% des simulations dans un cadre paramétrique donné satisfont ce critère, nous considérerons que le modèle est stable pour ce cadre paramétrique. Ce schéma d'analyse de l'interprétation confondante sera appelé « *Critère des γ* ».

4.3.5.3 Critère du score

Le score (la valeur estimée de la variable latente) n'est jamais discuté dans la littérature sur les modèles formatifs. Pourtant différents travaux émanant des pionniers des SEM (Jöreskog entre autres) considèrent que le recours aux scores dans certaines situations est pertinent (voir Bollen & Arminger, 1991; Jöreskog, Sörbom, & Wallentin, 2006; Raykov & Penev, 2013; Schumacker, 2002; Schumacker & Rigdon, 1995). De plus, à la différence d'une variable latente réflective, le score d'un modèle de mesure formatif est très aisé à estimer puisqu'il est par définition la combinaison linéaire de ses indicateurs (Bollen, 2011). Ainsi la comparaison du score estimé avec le score réel (« *true score* ») est un moyen de valider l'instabilité de la mesure formative sans se focaliser sur des indicateurs formatifs sujets à des instabilités liées à des colinéarités potentielles. Comme le vrai score est par définition connu en simulations Monte Carlo, il est aisé de calculer cette proximité entre le score véritable et son estimation, en prenant comme critère de proximité la corrélation (le score brut dépend du choix de normalisation).

Ici, nous considérons qu'il y a interprétation confondante lorsque la corrélation $\hat{\theta}_i$ entre le score estimé de la variable latente η (\hat{S}_f) et le score du modèle théorique S_f de référence dépasse un seuil.

Pour chaque simulation f nous calculons le score estimé (voir Bollen, 2011) :

$$\hat{S}_f = \sum_1^4 \hat{\lambda}_i \times X_i \quad (27)$$

La corrélation est estimée à partir du vrai score généré par la simulation :

$$\hat{\theta}_f = \text{Corr}(\hat{S}_f, S_f) \quad (28)$$

Le critère de proximité par la corrélation des deux latentes (empirique et théorique) sera abordé par le biais moyen (Hoogland & Boomsma, 1998).

$$B(\hat{\theta}) = \frac{\bar{\theta} - \theta}{\theta} \quad (29)$$

Avec ici $\theta = 1$ (corrélation théorique parfaite), nous considérons un biais $B(\hat{\theta}) = \frac{\bar{\theta} - \theta}{\theta} < 0.05\%$ comme une indication de l'absence d'interprétation confondante. Ce second critère sera appelée « *Critère d'identité* ».

Nous utiliserons ensuite un troisième critère, moins contraignant, basé sur le critère de redondance de Chin (2010). Chin considère qu'une corrélation $\hat{\theta}_i$ moyenne doit être supérieure à 0.80 (corrélation forte). Ce troisième critère sera appelé « *Critère de proximité* ».

Nous utilisons de nouveau une statistique de comptage : pour chaque configuration, nous retenons les simulations qui satisfont les conditions énumérées ci-dessus. Si 95% des simulations dans un cadre paramétrique donné satisfont le critère, nous considérerons que le modèle est stable pour ce cadre paramétrique.

4.3.5.4 Critère du lien structurel

Pour regarder l'efficacité des estimations du modèle structurel, nous regarderons le biais moyen relatif des λ , calculé comme la moyenne des deux biais en valeur absolue, relatif à λ_1 et λ_2 (Figure 32) :

$$B(\lambda) = \frac{\frac{|\bar{\lambda}_1 - \lambda_1|}{\lambda_1} + \frac{|\bar{\lambda}_2 - \lambda_2|}{\lambda_2}}{2} \quad (30)$$

Je précise que ce test n'a été fait que pour la phase 1 des simulations, donc uniquement avec le modèle 1 (deux latentes réfléchives). Ce critère est appelé « *critère des λ* ».

Là encore une statistique de comptage est employée : pour chaque configuration, nous retenons les simulations qui satisfont les conditions énumérés ci-dessus. Si 95% des simulations dans un cadre paramétrique donné satisfont le critère, nous considérerons que le modèle est stable pour ce cadre paramétrique.

4.3.5.5 Critère pour les contraintes de proportionnalité

Dans ce cadre nous analysons des modèles « mal spécifiés » et regardons si les MI peuvent détecter cela. Pour chaque cadre paramétrique nous regardons s'il y a au moins un MI supérieur à 7.88. Si tel est le cas, nous considérerons que la mauvaise spécification est détectée. Je rappelle que pour ces simulations, c'est le modèle MIMIC (modèle théorique : Figure 34 ; modèle estimé : Figure 31)

Une statistique de comptage est employée de nouveau avec toujours un seuil de 95% pour chaque cadre paramétrique.

4.3.5.6 Critère pour l'exhaustivité des indicateurs

Ici aussi le modèle théorique est mal spécifié. Nous calculons pour chaque simulation le \widehat{SMC} et le comparons au SMC théorique. Nous analysons le SCM moyen (\overline{SMC}) pour un cadre paramétrique pour voir s'il diverge du SMC théorique. Un seuil de 5% de biais sera pris, mais nous regarderons aussi si le \overline{SMC} est supérieur à 0.6 (le seuil minimum d'acceptabilité usuellement retenu). Ici de nouveau les estimations sont basées sur le modèle MIMIC (Figure 31).

4.3.6 Le code simulé

Cette étude adopte la définition d'un modèle formatif tel que défini par Bollen (2007, 2011) pour générer les matrices de covariances, ce qui correspond aux cadre de la littérature académique (Cadogan et al., 2008; Cenfetelli & Bassellier, 2009; Jarvis et al., 2003; Petter et al., 2007). Le code est donné en annexe, mais j'en explique ici les grandes étapes. Ce code souffre d'avoir été fait par une personne de la vieille école (j'ai été formé sur FORTRAN et PASCAL il y a 30 ans et me suis formé en autodidacte sur R), donc sans profiter de toutes les capacités de R en termes de gain de temps machine ; mais ceci permet de lire le code même sans être expert de R. De même le code des modèles CB-SEM est lourd à lire, mais il permet une lecture aisée des réels paramètres entrés. Le code avec LAVAAN aurait été beaucoup plus simple à écrire... mais LAVAAN donne des résultats qui semblent faux sur certains paramètres estimés. Le code a en effet été vérifié autant sur la genèse des données que sur l'efficacité des résultats en comparant nos résultats avec MPlus. Si nous avons pris le package *sem* au lieu de LAVAAN c'est que ce dernier ne donne pas des estimations forcément fiables des modèles formatifs, de plus les *Modifications d'Indices* (MI) sont

apparus fausses alors que *sem* donne des résultats quasi-identiques à ceux de *MPlus* sur l'ensemble des estimations pour tous les jeux de données testés.

Voici la structure du code (lisible en annexe) :

Le code génère en premier une matrice de corrélation (*SIGMA*) entre les indicateurs formatifs de la taille de l'échantillon (*taille*);

Nous générons ensuite une matrice *FORMAT* des indicateurs formatifs gaussiens centrés réduits via la fonction de R *rmvnorm* qui génère des données à partir d'une matrice de corrélation (ici *SIGMA*) ;

La latente formative (centrée réduite) est générée comme combinaison linéaire des indicateurs formatifs plus une erreur de sorte que la corrélation globale de la régression de la latente sur les indicateurs donne une corrélation désirée (*CORRO*) :

$$n_f = \gamma \times \sum_1^4 X_i + \zeta \quad (31)$$

avec $\zeta \sim N(0, \sigma)$ en spécifiant σ pour avoir la corrélation globale souhaitée. Concrètement le coefficient γ qui détermine le lien entre les indicateurs et la latente (identique pour tous les indicateurs formatifs) est déterminé par la corrélation car c'est le poids relatif qui nous intéresse.

Ce qui donne le code suivant :

$$LATENTE = scale(scale(rowSums(FORMAT)) + rnorm(taille, 0, \left(\frac{1-CORRO}{CORRO}\right)^{\frac{1}{2}}))$$

Remarque : la fonction *scale* centre et réduit des données et *rnorm* génère un vecteur gaussien de taille *taille*, de moyenne 0 et d'écart-type ici $\left(\frac{1-CORRO}{CORRO}\right)^{\frac{1}{2}}$ qui assure que la corrélation globale soit *CORRO* ;

Les latentes réflectives (*LAT1*, *LAT2*, ...) sont construites à partir de la latente formative en générant une erreur permettant d'obtenir une corrélation entre la latente formative et la latente réflective voulue (*cor1*, *cor2*, ...) de la même manière que pour la genèse de la latente formative :

$$LATENTE = scale(LAT + rnorm(taille, 0, \left(\frac{1-cor1}{cor1}\right)^{\frac{1}{2}}))$$

Remarque : le programme permettait de générer des corrélations différentes pour les différentes latentes réflectives (liens structuraux différenciés), mais pour simplifier les

résultats nous ne considérerons que des latentes réfléchives ayant une même corrélation avec la latente formative.

Les variables manifestes des latentes réfléchives sont générées comme précédemment avec une corrélation imposée à 0.9 :

$$Y1 = scale(LAT1 + rnorm(taille, 0, \left(\frac{1-0.9}{0.9}\right)^{\frac{1}{2}}))$$

Ce code permet de générer des données de variables « observables » (les indicateurs formatifs et les variables manifestes) qui respectent le cadre du modèle théorique posé. Par la suite nous estimons les paramètres par un modèle d'équations structurelles (soit le package *sem* pour CB-SEM, soit le package *plsem* pour PLS-SEM).

Avec CB-SEM la normalisation se fait, comme préconisé par (Franke et al., 2008), en mettant les variances des latentes à 1.

4.3.7 Les simulations

Le nombre de simulations par cadre paramétrique est de 500, comme le font la quasi-totalité des travaux de simulations Monte Carlo. Concrètement pour chaque cadre paramétrique, nous générons 500 jeux de données et estimons les paramètres. Afin d'éviter que des données qui ne permettraient pas au modèle de converger nuisent aux résultats, Chen et al. (2008) proposent de générer plus de modèles que nécessaire et d'en conserver 500 qui convergent ; il est apparu plus simple d'expurger les modèles non convergents et de faire un test sur le nombre de simulations ayant convergées.

Afin de s'assurer de l'efficacité de nos résultats, nous avons généré avec le modèle 1, 100 paires de latentes réfléchives pour une même latente formative (ce qui revient à faire ce que Bollen ou Howell ont fait dans leurs exemples). Pour un même cadre paramétrique nous avons effectué 100 simulations (soit 100 fois 100 paires générés), et comme prévu les résultats corroborent ceux obtenus par nos simulations présentées. Mais à la différence de Bollen ou Howell, il me paraît plus correct de générer des modèles théoriques et de voir si ceux-ci donnent des estimations valides, plutôt que de générer des paires de réfléchives. En effet, même si pour deux paires différentes les résultats des coefficients sont proches, rien ne dit que ces coefficients sont pour autant non-biaisés. C'est bien le biais potentiel que nous cherchons à estimer. Et si des coefficients varient fortement (interprétation confondante), cela se concrétisera par des biais importants au niveau des estimations.

4.3.8 Les résultats

J'ai généré au final des centaines de milliers de modèles différents. Il est impossible de présenter l'intégralité des résultats. Je me concentrerai sur les résultats les plus synthétiques et apportant des informations opérationnalisables. De même, il est impossible de rapporter non seulement l'ensemble des données générées, mais même impossible de donner des tableaux synthétiques par cadre paramétrique car ces résultats rempliraient des dizaines de pages.

Le temps machine est un peu long pour tout générer (mon code a demandé parfois plusieurs semaines de calculs à mes deux ordinateurs, sans discontinuité, pour certains cadres paramétriques), mais ce dernier est facilement utilisable (même par un non-expert de R) et le lecteur soucieux de vérifier mes résultats peut reprendre le code en annexe, l'analyser, et générer ses propres résultats.

4.4 L'interprétation confondante

L'interprétation confondante (explicitée dans la section antérieure) est au cœur de toute une discussion depuis 2007 (article de Howell et al., 2007b). Jusqu'à l'article de Bainter & Bollen (2014) la discussion n'avaient porté que sur quelques exemples (Franke et al., 2008; Howell et al., 2007b; Howell R.D., Breivik E., & Wilcox J.B., 2007; G. Kim et al., 2010; Wilcox et al., 2008)... critiquables (comme je l'ai expliqué dans la section précédente). De plus les critères retenus pour discuter de l'interprétation confondante me semblent eux aussi critiquables. J'utiliserai donc les critères proposés dans la section précédente pour mener la discussion via des simulations.

Le point central des discussions porte sur les sources potentielles de l'instabilité des estimations des paramètres du modèle de mesure, à savoir l'hypothèse de l'inhérence des mesures formatives à l'interprétation confondante (Howell et al., 2007b) ou l'hypothèse de la mauvaise spécification du modèle de mesure (Bollen, 2007). Ces auteurs ne discutent que de l'approche CB-SEM, je vais donc cadrer la discussion sur CB-SEM pour revenir plus tard sur PLS-SEM. Mais auparavant je repose le problème de l'interprétation confondante en introduction.

4.4.1 L'interprétation confondante

C'est Burt (1976) qui a été le premier à souligner l'importance de distinguer le sens nominal et le sens empirique d'un construit. Le sens nominal est celui postulé *à priori* par le praticien. Le sens empirique est obtenu *a posteriori* une fois le modèle de mesure qui représente le construit est estimé. Lorsqu'il y a divergence entre ces deux significations, les mesures spécifiées sont alors soumises à l'*interpretational confounding*. Qu'en est-il d'un construit représenté par des mesures formatives ?

Pour Howell et al. (2007a, 2007b); Wilcox et al. (2008), l'interprétation confondante est inhérente aux mesures formatives. Ils avancent plusieurs justifications. En premier, ces auteurs se réfèrent à l'hypothèse d'indépendance locale de la théorie classique du test (un construit existe indépendamment de ses manifestations), ils affirment que ce n'est pas le cas d'un construit appréhendé avec des mesures formatives. Comme ces dernières déterminent entièrement le construit, ce type de construit n'a pas d'existence réelle pour eux. J'ai abordé cette discussion ontologique en introduction. Il me semble compliqué de suivre ces auteurs car dans un article de 2008 ils écrivent « *Constructs themselves, posited under a realist*

philosophy of science as existing apart from their measurement, are neither formative nor reflective » (Wilcox et al., 2008, p. 1220). Et en effet du point de vue empiriste-réaliste un construit peut être considéré comme mesurable formativement.

Un autre argument plus important pour ces auteurs est que cette conceptualisation formative conduit par définition à une sous-identification du modèle de mesure ce qui nécessite alors l'ajout de variables latentes endogènes et/ou de mesures réfléchives pour son estimation (Figure 31 ou Figure 32). Pour ces auteurs, puisque l'identification exige des variables endogènes réfléchives, il n'est donc pas surprenant d'observer une dépendance des paramètres estimés du modèle formatif (paramètres γ , Figure 31 ou Figure 32) en fonction des variables endogènes réfléchives spécifiées (paramètres λ , Figure 31 ou Figure 32). Ainsi cette instabilité des paramètres γ du modèle formatif conduit inéluctablement à l'ambiguïté d'interprétation de la variable latente η (Figure 31 ou Figure 32) car le sens empirique change d'un SEM à l'autre selon les variables endogènes introduites. Le sens empirique divergerait ainsi de celui défini *a priori* (sens nominal).

Enfin, ces auteurs font remarquer que le fait que les indicateurs formatifs ne soient pas forcément corrélés, signifie qu'ils ne sont pas reliés au même réseau nomologique. N'ayant pas les mêmes antécédents et conséquences, il n'est pas alors étonnant pour Howell et al. (2007b) d'observer les valeurs des paramètres du modèle de mesure formatif varier pour s'adapter au SEM spécifié. Ce qui fragilise la validité externe d'un modèle formatif et donc son utilisation dans les applications empiriques à des fins de comparaisons et de généralisations. Depuis d'autres auteurs ont illustré ce problème et tiré les mêmes conclusions que Howell et al. (Franke et al., 2008; G. Kim et al., 2010; Wilcox et al., 2008)... avec des lacunes dans leurs données comme déjà dit.

Comme le souligne (Bollen, 2007), les exemples numériques reportés par Howell et al. (2007b) pour appuyer leur thèse ne permettent pas de se prononcer sur le bien-fondé de leur affirmation, à savoir que les mesures formatives sont intrinsèquement soumises à l'interprétation confondante. Bollen (2007), à juste titre, soulève une faiblesse essentielle dans la méthodologie de Howell et al. (2007b), à savoir l'absence du modèle théorique qui a permis de générer les matrices des variances-covariances entre les différentes mesures utilisées. Cet argument est d'autant plus convaincant que Bollen (2007) montre sur des données simulées, à partir d'un modèle théorique comportant un modèle de mesure formatif d'une latente exogène et d'autres mesures endogènes réfléchives, l'absence d'interprétation confondante. Des simulations confirment qu'un modèle formatif n'est pas intrinsèquement

instable (Bainter & Bollen, 2014)... même si comme déjà dit le spectre des paramètres utilisés est en réalité très étroit. Ainsi pour Bollen (2007) l'interprétation confondante mise en évidence par Howell et al. (2007b) n'est que la conséquence d'une mauvaise spécification du modèle de mesure formatif. J'ai montré en effet que les exemples de Howell et al. étaient mal spécifiés, ce qui va dans le sens de Bollen. D'ailleurs, cette critique n'est pas rejetée par Howell et al. (2007a). Mais ils soulignent la difficulté de détecter une mauvaise spécification du modèle de mesure sur des données réelles car le modèle de mesure formatif ne peut être testé isolément (problème d'identification).

4.4.2 Simulations

Nous partons d'un modèle bien spécifié. J'amorce par ce que j'ai nommé la phase 2 des simulations (au sens chronologique). Deux configurations sont retenues : Modèle 1 avec (deux latentes réflectives) et modèle 2 (avec quatre latentes réflectives).

Comme dit, la normalisation se fait par la contrainte des variances des latentes à 1.

Le test sur les coefficients se fait par des tests LR avec un contrôle de Bonferonni.

Les paramètres utilisés pour la phase 2 sont :

- Taille de l'échantillon : 250, 500, 1000 (afin de reprendre le cadre de (Bainter & Bollen, 2014))
- Le SMC variera de 0.5 à 0.9 par pas de 0.1
- Les coefficients structureaux λ sont tous identiques et varieront de 0.2 à 0.9 par pas de 0.1
- Les variables manifestes des latentes formatives sont imposées avec une corrélation identique à 0.9

Le nombre de configurations présentées dans cette phase 2 de simulations est donc 288 ($= 2 \times 3 \times 6 \times 8$) ; 2 modèles, 3 tailles d'échantillon, six paramètres pour le SMC, 7 paramètres pour les λ . Chaque configuration est simulée 500 fois, soit 144 000 modèles. Dans ce travail, afin de ne pas être pollué par la colinéarité des coefficients formatifs, nous avons contraint ces coefficients à être non corrélés.

Une première phase de simulations (phase 1) avait été réalisée sur le modèle 1, donc sans prendre en compte un modèle avec 4 réflectives endogènes (modèle 2), mais en intégrant la colinéarité des indicateurs formatifs. La stabilité des indicateurs formatifs a été faite sur la base des intervalles de confiance des indicateurs formatifs et non sur la base de tests LR, ces simulations ne sont donc pas efficaces. Cependant nous donnerons certains résultats à partir

de cette phase 1 des simulations. De plus, l'analyse sur le modèle structurel n'a été effectuée que pour la première phase de simulations, donc je m'appuierai dessus lors de la discussion sur ce problème.

Les paramètres de la phase 1 sont :

- N (Taille de l'échantillon) : 200, 300, 500, 750, 1000
- Le SMC variera de 0.6 à 0.9 par pas de 0.1
- Les coefficients structureux λ sont tous identiques et varieront de 0.2 à 0.8 par pas de 0.2 ;
- R : La colinéarité entre les indicateurs formatifs varie de 0 à 0.8 par pas de 0.2.

Soit pour cette phase 1 de simulations nous avons 400 configurations ($= 5 \times 4 \times 4 \times 5$) ; 5 tailles d'échantillon, 4 paramètres pour le SMC, 4 paramètres pour les λ , 5 paramètres pour R. Chaque configuration est simulée 500 fois, soit 200 000 modèles.

Comme déjà dit, les modèles qui ne convergent pas sont éliminés afin de s'assurer une réplication de 500 pour chaque cadre paramétrique.

Les résultats seront analysés en considérant N : taille de l'échantillon, SMC : la corrélation des indicateurs formatifs sur la latente formative, λ : les liens structureaux, R : colinéarité entre les indicateurs formatifs.

4.4.3 Résultats sur l'instabilité des paramètres formatifs

La phase 1 de simulations intégrait la colinéarité des indicateurs formatifs, mais le critère de stabilité est mesuré via des intervalles de confiance. Donc cette procédure peut être considérée comme potentiellement erronée, mais cela permet cependant de voir l'impact relatif des différents paramètres. Une analyse de variance est faite sur l'ensemble des simulations de cette phase 1, c'est-à-dire 200 000 résultats différents ($500 \times 5 \times 5 \times 4 \times 4$). Le Tableau 7 montre que le facteur le plus important est le vecteur des liens structurels λ . Pour les facteurs N et R les résultats obtenus vont dans le sens attendu : une taille d'échantillon importante améliore la qualité des résultats ; une colinéarité importante entre les indicateurs formatifs dégrade les résultats. Par contre le pourcentage expliqué de la variance de la variable latente (SMC) n'a pas d'impact sur la qualité d'ajustement. La dégradation consécutive de la colinéarité confirme comme attendu la nécessité de ne travailler qu'avec des modèles où les indicateurs formatifs sont non-corrélés.

Tableau 7 : Résultats de l'ANOVA pour le « critère des γ »

Variable	Coefficient (standardisé)	Significativité
N	0.287	<0.0001
R	-0.385	<0.0001
SMC	0.030	0.087
λ	0.709	<0.0001

Rsquare=0.75

Je ne m'étends pas plus sur cette phase 1 de simulations et considère que les simulations de la seconde phase, dont les paramètres ont été exposés dans la section précédente, donnent des résultats fiables pour le « critère des γ » car basés sur le critère LR. Pour rappel il y a 288 configurations retenues dans la phase 2 et j'utilise une statistique de comptage: le nombre de simulations (sur les 500) qui remplissent les conditions énumérés plus haut. Nous considérons pour une configuration donnée, qu'il y a absence d'une interprétation confondante lorsque au moins 95 % des simulations de cette configuration remplissent le « critère des γ », soit (475/500).

Les tableaux suivant donnent les liens structurels λ minimum pour avoir une stabilité des coefficients formatifs.

TABLEAU 8 : MODELE 1, LIEN STRUCTUREL MINIMUM POUR OBTENIR DES INDICATEURS FORMATIFS NON-INSTABLES

Model 1	SMC=0.4	SMC=0.5	SMC=0.6	SMC=0.7	SMC=0.8	SMC=0.9
N=250	$\lambda \geq 0.8$	$\lambda \geq 0.7$	$\lambda \geq 0.7$	$\lambda \geq 0.7$	$\lambda \geq 0.7$	$\lambda \geq 0.8$
N=500	$\lambda \geq 0.6$	$\lambda \geq 0.5$	$\lambda \geq 0.5$	$\lambda \geq 0.6$	$\lambda \geq 0.7$	$\lambda \geq 0.8$
N=1 000	$\lambda \geq 0.4$	$\lambda \geq 0.4$	$\lambda \geq 0.5$	$\lambda \geq 0.5$	$\lambda \geq 0.6$	$\lambda \geq 0.8$

TABLEAU 9: MODELE 2, LIEN STRUCTUREL MINIMUM POUR OBTENIR DES INDICATEURS FORMATIFS NON-INSTABLES

Model 2	SMC=0.4	SMC=0.5	SMC=0.6	SMC=0.7	SMC=0.8	SMC=0.9
N=250	$\lambda \geq 0.7$	$\lambda \geq 0.6$	$\lambda \geq 0.5$	$\lambda \geq 0.5$	$\lambda \geq 0.6$	$\lambda \geq 0.6$
N=500	$\lambda \geq 0.4$	$\lambda \geq 0.4$	$\lambda \geq 0.4$	$\lambda \geq 0.4$	$\lambda \geq 0.5$	$\lambda \geq 0.7$
N=1 000	$\lambda \geq 0.3$	$\lambda \geq 0.3$	$\lambda \geq 0.4$	$\lambda \geq 0.4$	$\lambda \geq 0.4$	$\lambda \geq 0.6$

Nos résultats montrent que la stabilité des paramètres dépend du nombre de variables latentes endogènes. Si le modèle structurel n'est pas trop faible (les λ), avec un échantillon important nous pouvons considérer avoir des estimations fiables de nos indicateurs formatifs. Pour des tailles d'échantillons plus faibles (250 ou 500), il est nécessaire de s'assurer que les coefficients structurels soient plus élevés.

Le SMC a un rôle neutre excepté lorsqu'il est très important (0.9). De prime abord un tel SMC doit être considéré comme rarissime, de plus cet impact est certainement conséquent de l'algorithme qui cherche à minimiser les variances résiduelles des latentes réflectives, entraînant des biais des indicateurs formatifs... ce que Howell. et al. pointaient comme argument.

Mais en dehors d'un SMC très élevé, les coefficients formatifs paraissent stables à condition d'avoir des liens structurels relativement importants. La stabilité des indicateurs formatifs dépend bien des variables latentes réflectives introduites, et l'instabilité semble potentielle avec un modèle n'intégrant que deux latentes réflectives. Ce résultat confirme donc partiellement ce que Howell et al. (2007b) qualifiaient de dépendance des estimations des paramètres du modèle de mesure aux variables endogènes retenues dans la SEM, et donc d'inhérence des mesures formatives à une interprétation confondante. Néanmoins lorsque les paramètres λ modèle structurel ne sont pas trop faibles, nous n'observons pas d'interprétation confondante. En d'autres mots, le modèle structurel impacte les paramètres du modèle formatif, mais il n'y a pas une instabilité inhérente à tous modèles formatifs ; ce qui confirme les propos de Bainter & Bollen (2014, 2015).

4.4.4 Résultats sur le score

Nous examinons l'absence de l'interprétation confondante sur la base des corrélations entre les scores estimés de la latente et son score théorique, selon deux critères (décrits précédemment) : 1/ le « *critère d'identité* » ; 2/ le « *critère de proximité* ». Pour rappel un biais important indique l'existence d'une interprétation confondante.

Nous procédons de prime abord de la même manière que précédemment, par une analyse de la variance sur la phase 1 de simulations ; avec ici comme variable dépendante $\hat{\theta}_i$ (la corrélation entre le score estimé $\hat{\eta}$ et le vrai score η) et comme variables indépendantes les facteurs N, R, SMC, et λ ; et ceci sur la base des 400 000 simulations.

TABEAU 10 : RESULTATS DE L'ANOVA SUR LA CORRELATION ENTRE LE SCORE ESTIME $\hat{\eta}$ ET LE VRAI SCORE η

Variable	Coefficient (standardisé)	<i>p-value</i>
N	0.202	<0.0001
R	0.00029	0.98
SMC	0.524	<0.0001
λ	0.729	<0.0001

Rsquare = 0.85

Il est intéressant de noter qu'à la différence de l'analyse basée sur « *critère des γ* », ici le coefficient R ne joue aucun rôle direct sur le niveau du fit du SEM. Cela confirme en premier lieu que le critère de stabilité des indicateurs formatifs est un critère erroné car dans tous les exemples proposés par les différents auteurs il y avait une corrélation entre les indicateurs formatifs. On ne peut analyser les coefficients formatifs indépendamment de leur colinéarité. Par contre, le SMC a une réelle influence sur la qualité d'ajustement du SEM, ce qui là aussi va dans le sens attendu. Le vecteur des liens structurels λ reste le facteur le plus influent.

4.4.4.1 Critère d'identité

Je reprends les modélisations telles que exposées précédemment dans la phase 2 afin d'avoir des résultats en ligne avec ceux du test de stabilité des indicateurs formatifs. Nous retenons

dans un premier temps le « *critère d'identité* », c'est-à-dire le critère tel que le biais de la corrélation moyenne estimée soit inférieure à 5% (à savoir : $B(\hat{\theta}) = \frac{\bar{\theta} - \theta}{\theta} \leq 0.05$).

Les résultats montrent que, quel que soit le niveau des paramètres, aucune configuration ne montre une identité du score estimé avec le score théorique.

4.4.4.2 Critère de proximité

Ce seuil de corrélation moyenne supérieure à 0.95 peut être considéré comme trop strict. En effet, la corrélation moyenne sur toutes nos simulations vaut 0.87 (médiane=0.89) avec un écart type de 0.0925 (la corrélation minimum est à -0.45) ; notons que nous ne retenons ici que les simulations ayant convergé avec CB-SEM. Ce résultat indique qu'en général, lorsque le SEM converge avec CB-SEM, le score estimé de la latente est fortement corrélé au score théorique. Nous utilisons donc un autre critère plus souple, « *critère de proximité* », qui consiste à calculer le nombre de corrélations (sur 500 simulations) supérieur à 0.8 (critère de *redondance* de Chin (2010)). Nous exigeons une proportion minimale de 0.95 (95% au moins des 500 simulations donnent une corrélation supérieure à 0.80 entre les scores théorique et estimé). Les Tableau 11 et Tableau 12 rapportent les valeurs minimales de N, du SMC, et les λ nécessaire pour obtenir la proportion de 95% de corrélations supérieure à 0.8.

TABLEAU 11 : MODELE 1, LIMITE INFÉRIEURE DES λ POUR VALIDER LE CRITÈRE DE « PROXIMITÉ »

$\lambda \geq$	SMC = .4	SMC = .5	SMC = .6	SMC = .7	SMC = .8	SMC = .9
N = 250	/	/	/	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .9$
N = 500	/	/	/	$\lambda \geq .6$	$\lambda \geq .7$	$\lambda \geq .8$
N = 1000	/	/	/	$\lambda \geq .5$	$\lambda \geq .6$	$\lambda \geq .7$

TABLEAU 12 : MODELE 2, LIMITE INFERIEURE DES λ POUR VALIDER LE CRITERE DE « PROXIMITE »

$\lambda \geq$	SMC = .4	SMC = .5	SMC = .6	SMC = .7	SMC = .8	SMC = .9
N = 250	/	/	/	$\lambda \geq .4$	$\lambda \geq .3$	$\lambda \geq .2$
N = 500	/	/	/	$\lambda \geq .3$	$\lambda \geq .2$	$\lambda \geq .2$
N = 1000	/	/	/	$\lambda \geq .2$	$\lambda \geq .2$	$\lambda \geq .2$

Avec deux latentes réflectives, il faut un SMC supérieur à 0.6 et des liens structuraux « medium » pour assurer une proximité entre le score théorique et le score estimé. Avec 4 réflectives, le seuil des liens structuraux peut être abaissé à un niveau relativement faible.

Au bilan, un modèle formatif correctement spécifié (un SMC > 0.6) couplé à un modèle structurel avec des liens pas trop faibles, souffre peu d'interprétation confondante au sens du « *critère de proximité* ». Ainsi, lorsque le SEM est correctement spécifié le score estimé de la latente converge vers le score théorique.

4.4.5 Validité du modèle structurel

J'ai souligné l'importance des paramètres λ du modèle structurel (c.à.d. le « *critère des λ* »). Nous avons conduit des simulations pour évaluer l'ampleur des biais d'estimation de ces paramètres (équations (30)). Ces simulations (phase 1) n'ont été faites qu'avec le modèle 1 (deux latentes réflectives) ; et donc nous ne pouvons estimer l'impact du nombre de variables endogènes.

Nos résultats donnent un biais moyen relatif à 0.1305. Ce biais relatif diminue dès que λ ou N croît, mais augmente dès que le SMC > 0.8 (cf. tableau 7 pour des résultats synthétiques).

TABLEAU 13 : MODELE 1, LIMITE INFERIEURE DES λ POUR VALIDER LE MODELE STRUCTUREL

$\lambda \geq$	SMC=0.6	SMC=0.7	SMC=0.8	SMC=0.9
N=200	0.6	0.6	0.6	0.6
N=300	0.6	0.6	0.6	0.6
N=500	0.4	0.4	0.6	0.4
N=750	0.4	0.4	0.4	0.4
N=1000	0.4	0.4	0.4	0.4

Avec $N=300$, il faut $\lambda \geq 0.6$ pour avoir un biais moyen relatif inférieur à 5% ; avec $N \geq 500$, pour avoir un biais moyen relatif inférieur à 5% il suffit d'avoir $\lambda \geq 0.4$. En conclusion, les liens structuraux semblent peu biaisés... autant que ces liens structuraux soient élevés. Il y a donc une circularité. Mais le biais absolu le plus important est de 0.15 et ne dépasse jamais 0.10 pour $N \geq 300$. Concrètement, un praticien ayant $N \geq 300$ pourra considérer que ses estimations de λ ont un biais absolu inférieur à 0.10. Donc il faudra s'assurer avoir des $\hat{\lambda}$ au moins à 0.70 pour s'assurer des λ au moins à 0.60 pour des échantillons de tailles inférieures à 500 ; et des $\hat{\lambda}$ au moins à 0.50 pour des tailles d'échantillons supérieurs à 300.

Nous retrouvons un cadre paramétrique proche de ceux de la fiabilité de la mesure d'un modèle formatif discuté auparavant. De nouveau donc les liens structuraux sont centraux pour s'assurer un modèle structurel fiable en termes d'estimation, mais à un certain seuil les estimations semblent peu biaisées.

4.4.6 Conclusion sur le modèle de mesure formatif

Nos résultats montrent d'une part que le critère retenu par la littérature pour discuter des modèles de mesure formatifs (i.e. la stabilité des indicateurs formatifs) n'est certainement pas le plus pertinent car la valeur de ces estimations dépend de la colinéarité des indicateurs formatifs. Le test sur les scores est certainement plus efficace pour s'assurer la stabilité et la fiabilité du modèle de mesure.

Nos résultats indiquent que dans un certain cadre paramétrique un modèle de mesure formatif peut donner des estimations non biaisées autant sur le modèle mesure que sur les liens structuraux. Le nombre de variables réflexives endogènes influence la fiabilité du modèle de mesure. Avec seulement deux latentes réflexives, les coefficients formatifs, le score estimé et le modèle structurel seront biaisés si les liens structuraux sont inférieurs à un seuil

« medium »... rarement trouvé dans la littérature académique. En effet, je n'ai pas connaissance d'articles sur des modèles de mesures formatifs reliés à uniquement deux latentes réflectives exhibant des liens structuraux standardisés supérieurs à 0.6.

Avec un modèle avec 4 réflectives endogènes, les critères peuvent être relâchés. Les liens structuraux nécessaires pour avoir une estimation fiable des indicateurs formatifs et du score estimé apparaissent en phase avec ce que la littérature peut montrer comme résultats empiriques. Ce cadre paramétrique exige un SMC important (le seuil de 0.6 usuellement retenu semble opérationnel) et des liens structuraux plus ou moins importants selon la taille de l'échantillon.

Ces résultats nuancent ce que certains auteurs, critiques des modèles de mesure formatifs, ont considéré : pour eux les modèles formatifs sont, de façon inhérente, sujet à une instabilité. Tous nos résultats semblent montrer que dans un cadre paramétrique précis le modèle de mesure formatif donne des résultats peu biaisés, autant sur les indicateurs formatifs (s'il n'y a pas de colinéarité entre eux) que sur le score estimé ou sur les liens structuraux.

Cependant nos simulations confirment l'importance des liens structuraux, c'est-à-dire des liens avec les variables endogènes qui servent pour l'identification du modèle. Le rôle des liens structuraux peut se considérer comme inscrit dans le cadre de la bonne spécification du modèle (validité convergente) comme le disent Bainter & Bollen (2015). Cependant, la dépendance du modèle de mesure à des variables endogènes va dans le sens de la critique de Rhemtulla et al. (2015) qui considère alors qu'un modèle de mesure formatif est en réalité un modèle de mesure réflectif. En effet, s'il faut des liens structuraux importants pour considérer comme fiable le modèle de mesure, nous sommes concrètement dans le cadre d'une mesure réflective de second ordre. Je vais revenir sur ce problème par simulations dans la section suivante.

4.5 Une latente formative est-elle réflexive ?

J'aborde maintenant une autre critique des modèles formatifs : ne sont-ils pas en réalité des modèles de mesure réflexifs ? Les seuls arguments sont d'ordres théoriques et aucune simulation ne semble avoir été effectuée pour cette question. Je vais donc présenter des simulations pour discuter de cette question.

4.5.1 Les arguments théoriques

Howell (2014); Lee et al. (2013); Rhemtulla et al. (2015) ont récemment repris l'argument contre les modèles formatifs en considérant que, puisqu'il fallait des liens structuraux pour identifier le modèle formatif et de plus que ces liens soient importants pour éviter les biais de mesure, un modèle de mesure formatif est en réalité un modèle de mesure réflexif.

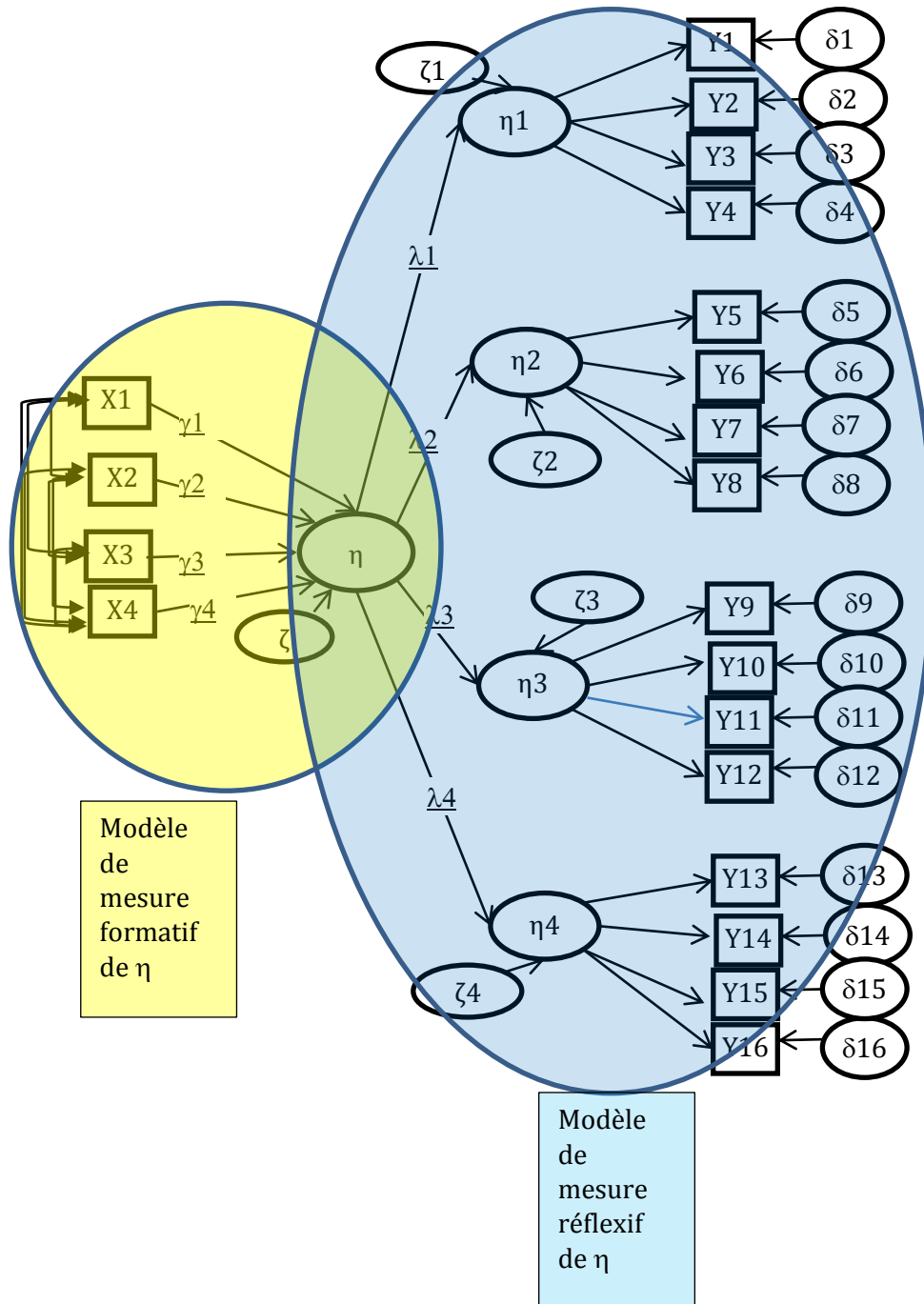
Pour ces auteurs, le modèle de mesure formatif (les indicateurs formatifs) n'apporte aucune information sur le sens de la latente. La variable latente est considérée donc à tort comme formative car les indicateurs formatifs sont « *superfluous* » (Rhemtulla et al., 2015) pour donner le sens de la latente considérée car tout le sens provient des mesures réflexives (les latentes réflexives de premières ordres). Pour ces auteurs, les indicateurs formatifs ne donnent aucune information supplémentaire.

Si cet argument paraît pertinent, aucune simulation ne semble jamais avoir été faite pour valider cette hypothèse. Je vais donc rapporter par la suite des simulations qui discutent de ce problème.

4.5.1.1 Les deux modèles de mesure en concurrence

Le modèle MIMIC (Figure 31), comme expliqué en introduction, peut être assimilé à un modèle réflexif. Mais en est-il de même avec un modèle intégrant des latentes réflexives comme le montre la Figure 35.

FIGURE 35 : MODELE DE MESURE FORMATIFS VS MODELE DE MESURE REFLECTIF D'UNE MEME LATENTE



Comme illustré dans la Figure 35, avec des coefficients élevés (λ_i), la latente η peut être vue comme une latente réflexive de second ordre avec quatre latentes réflexives de premier ordre η_1, η_2, η_3 et η_4 . En conséquence le sens de η semble entièrement donné par ces latentes réflexives. Le modèle de mesure serait donc :

$$\eta_i = \lambda_i \eta + \zeta_i \quad (32)$$

Et le modèle formatif (équation (33)) serait superflu.

$$\eta = \sum_i \gamma_i \times X_i + \zeta \quad (33)$$

A l'inverse, pour Bainter & Bollen (2014, 2015) l'équation (32) est un modèle structurel et non un modèle de mesure : la latente η est une variable prédictive de η_1 , η_2 , η_3 et η_4 . Pour ces auteurs le modèle de mesure est formatif (équation (33)). Il y a ainsi deux conceptions opposées.

4.5.1.2 La discussion analytique

Si aucune simulation n'a été faite, cet argument sur le caractère superflu du modèle formatif a été développé analytiquement par Howell et al. (2007a). A partir d'un modèle MIMIC

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \zeta \quad (34)$$

$$y_1 = \lambda_1 \eta + \varepsilon_1 \quad (35)$$

$$y_2 = \lambda_2 \eta + \varepsilon_2 \quad (36)$$

Franke et al. (2008); Howell et al. (2007b); Wilcox et al. (2008) affirment que les corrélations entre les variables réflectives impactent mécaniquement la variance de l'erreur ζ liée à la latente formative (η), c'est à dire le SMC estimé (Howell et al., 2007a) : "*When y_1 and y_2 correlate strongly, there is more covariation (as expressed in η) to be explained, and thus more error variance, and vice versa*"⁴⁴. We note that this challenges Bollen's (2007) definition of ζ : "*a term that includes all other determinants of η that are not included in the equation*" (p. 220)" » (p. 242). (Wilcox et al., 2008): « *The error term then, is realized as the shared variance between the outcomes*⁴ *not accounted for by the formative indicators. Thus, surplus meaning has to do with the endogenous variables, not the formative construct*⁴ [...] *the interpretation of the disturbance term as being associated with the formative construct appears to be ill-founded.* » (p. 1224-1225) and « *The meaning of the error term associated with formatively measured constructs in structural equations models is more closely associated with the constructs dependent*⁴ *on the formative construct and their correlation than on the formative measures* » (p. 1227).

Ce lien entre les variances des λ et de l'erreur de la variable latente η me paraît peu fondé. Une forte corrélation entre les variables réflectives endogènes n'entraînera pas forcément une

⁴⁴ Souligné par moi

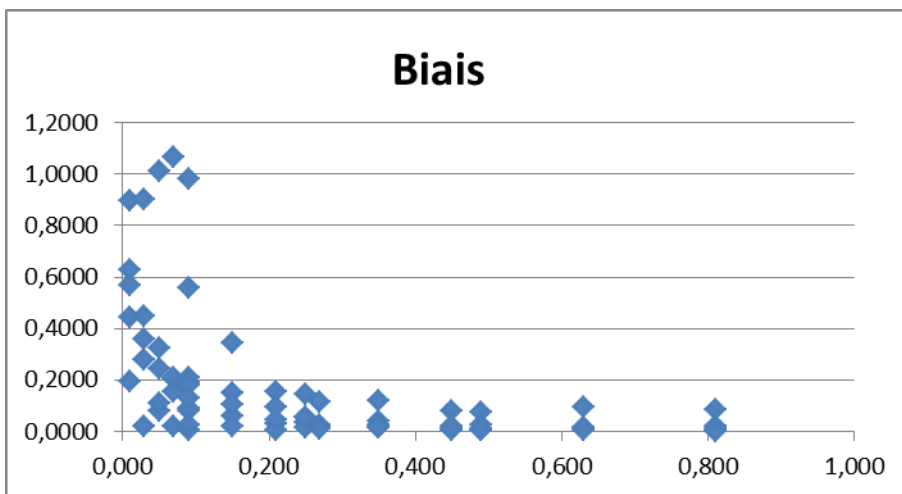
augmentation de la variance de l’erreur de η , au contraire même car plus cette corrélation sera forte plus le modèle de mesure MIMIC sera valide en raison du principe d’interchangeabilité des mesures réflectives, comme le montre l’équation suivante (cas du modèle MIMIC, Figure 31):

$$Cov(Y_1; Y_2) = \lambda_1 \lambda_2 VAR(\eta) = cov(\lambda_1(\gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3) + \lambda_1 \zeta + \varepsilon_1; \lambda_2(\gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3) + \lambda_2 \zeta + \varepsilon_2) = \lambda_1 \lambda_2 VAR(\gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3) + \lambda_1 \lambda_2 VAR(\zeta) \quad (37)$$

Même avec une corrélation élevée entre Y1 et Y2, la variance de l’erreur ζ de la variable formative η sera faible lorsque la variable latente formative η est entièrement déterminée par ses indicateurs formatifs (hypothèse d’exhaustivité). Pour appuyer nos arguments, nous avons généré des SEM avec un échantillon de taille N= 500, en posant différents SMC pour la latente formative (0.1 ; 0.3 ; 0.5 ; 0.7 ; 0.9), et en posant différentes corrélations entre η et Y1 et entre η et Y2, corrélations identiques allant de 0.1 ; 0.3 ; 0.5 ; 0.7 ; 0.9, ce qui donne une corrélation $\rho(Y1 ; Y2)$ allant de 0.01 à 0.81. Pour chaque configuration nous avons effectué 1000 répliques.

Notre objectif est de mesurer l’ampleur du biais potentiel entre le SMC théorique et le SMC estimé, conséquent de la variation de la corrélation entre Y1 et Y2. Les résultats tendent à infirmer les hypothèses de Howell et al. (2007b). On remarque (Figure 36) que le biais du SMC estimé est très faible dès que le SMC théorique et la corrélation théorique entre les endogènes dépassent le seuil de 0.2. De même, on remarque que la variance de l’erreur est inversement proportionnelle au degré de la mauvaise spécification du SEM, et non en fonction de la corrélation entre les mesures réflectives soutenues par Howell et al. (2007b)

FIGURE 36 : BIAIS DE \widehat{SMC} EN FONCTION DES VALEURS DES PARAMETRES $\lambda_1 * \lambda_2$



Ces résultats empiriques confirment que Howell et al. ont tort dans leur argument et confirme ce que Bollen considère de la variance résiduelle de la latente formative : « *a term that includes all other determinants of η_{1i} that are not included in the equation* » (Bollen, 2007, p. 220). En conséquence le sens de la variance de l'erreur de la latente n'est pas consécutive des réflectives mais dû à la mauvaise spécification du modèle, en conséquence une erreur importante indique « *that the full model is misspecified* » (Bainter & Bollen, 2014, p. 135). Donc sur le plan analytique, rien ne confirme l'hypothèse du caractère superflu des indicateurs formatifs pour donner le sens de la latente.

4.5.2 Simulations

L'objectif est de vérifier par simulations l'hypothèse que le modèle de mesure de la latente η de la Figure 32 ou de la Figure 33 peut être assimilé à une latente réflective (modèle de mesure réflectif de la Figure 35). Nous partons du modèle 1 et 2 dans un premier temps, puis seul le modèle 2 sera utilisé par la suite pour des problèmes d'identification pour certaines simulations.

4.5.2.1 Paramètres des simulations

Les paramètres sont ceux de la phase 2 des simulations :

- Taille de l'échantillon : 250, 500, 1000 (afin de reprendre le cadre de (Bainter & Bollen, 2014))
- Le SMC variera de 0.5 à 0.9 par pas de 0.1
- Les coefficients structuraux λ sont tous identiques et varieront de 0.2 à 0.9 par pas de 0.1
- Les variables manifestes des latentes formatives sont imposées avec une corrélation identique à 0.9

Le nombre de configurations présentés dans cette phase 2 de simulations est donc 288 ($=2 \times 3 \times 6 \times 8$) ; chaque configuration est simulée 500 fois, soit 144 000 modèles.

4.5.2.2 Critères sur le score

De même que nous pouvons estimer la corrélation entre le score formatif (\hat{S}) et le vrai score S , nous pouvons calculer cette corrélation à partir du score estimé du modèle de mesure réflectif (\hat{S}^*), en utilisant le score estimé par le logiciel utilisé (package *sem* de R).

Nous estimerons le biais moyen du score réflectif pour voir s'il représente une bonne estimation du score en prenant les critères « d'identité » et de « proximité » utilisés avec le score formatif. Puis dans un second temps nous comparerons les corrélations, si la corrélation entre \hat{S}^* et S est supérieure à la corrélation entre \hat{S} et S , le modèle de mesure formatif pourra être considéré comme superflu.

4.5.2.3 Critère par un test LR

Nous pouvons aborder l'hypothèse du caractère superflu du modèle formatif par un test LR. Nous comparons le modèle global (Figure 33) et le modèle contraint (modèle réflectif de la Figure 35, c'est-à-dire expurgé des indicateurs formatifs). Si le test rejette H_0 , cela voudra dire que le modèle formatif n'est pas statistiquement superflu et les indicateurs formatifs comme apportant de l'information significative (au sens statistique).

Notons que ce test ne peut être fait que sur le modèle 2 avec 4 latentes réflectives car le modèle avec seulement deux latentes réflectives n'est pas estimable sans les indicateurs formatifs.

4.5.3 Résultats

Je donne en premier lieu les résultats sur les corrélations puis j'analyse les résultats du test LR avant de conclure

4.5.3.1 Résultats sur les corrélations

Avec le critère « d'identité », seuls les modèles ayant des liens structuraux de 0.9 assurent une identité entre le score réflectif estimé et le score théorique. Ce résultat n'est pas surprenant car avec de tels liens structuraux il est difficile de ne pas considérer le modèle de mesure comme réflectif.

Mais dès que les liens structuraux diminuent, même au seuil de 0.8 et même avec un échantillon de 1000 le modèle de mesure réflectif ne donne pas un score estimé identique au score théorique.

Si nous prenons le critère plus lâche de « proximité » (corrélation supérieure à 0.80 entre le score réflectif estimé et le score théorique), nous obtenons :

TABLEAU 14 : MODELE 1, LIMITE INFERIEURE DES λ POUR VALIDER LE CRITERE DE « PROXIMITE » POUR LA MESURE REFLECTIVE

$\lambda \geq$	SMC = .4	SMC = .5	SMC = .6	SMC = .7	SMC = .8	SMC = .9
N = 250	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$
N = 500	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$
N = 1000	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$

TABLEAU 15: MODELE 2, LIMITE INFERIEURE DES λ POUR VALIDER LE CRITERE DE « PROXIMITE » POUR LA MESURE REFLECTIVE

$\lambda \geq$	SMC = .4	SMC = .5	SMC = .6	SMC = .7	SMC = .8	SMC = .9
N = 250	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$
N = 500	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .7$
N = 1000	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$	$\lambda \geq .6$

Les résultats dépendent légèrement de la taille de l'échantillon, mais surtout des liens structuraux. Il faut des liens structuraux importants pour pouvoir avoir une estimation fiable de la mesure réflective. Le seuil de 0.7 trouvé ici corrobore le seuil usuel considéré comme acceptable pour un modèle de mesure réflectif.

Afin de comparer les deux modèles (formatif vs réflectif) je donne le meilleur modèle au sens d'une corrélation moyenne supérieure. A partir des moyennes et écart-type des estimations par cadre paramétrique, un test de Student est effectué :

TABLEAU 16 : MEILLEUR MODELE DE MESURE

$\lambda \geq$	SMC = .4	SMC = .5	SMC = .6	SMC = .7	SMC = .8	SMC = .9
MMF	$\lambda \leq 0.3(*)$	$\lambda \leq 0.4$	$\lambda \leq 0.5$	$\lambda \leq 0.6$	$\lambda \leq 0.7$	$\lambda \leq 0.8$
MMR	$\lambda \geq .4(*)$	$\lambda \geq .5$	$\lambda \geq .6$	$\lambda \geq .7$	$F\lambda \geq .8$	$\lambda \geq .9$

MMF : Modèle de mesure formatif; MMR : Modèle de mesure réflectif

(*) pour un échantillon de 250, il n'y a pas de différence significative entre les deux modèles de mesure pour $\lambda = .4$

Quelle que soit la taille de l'échantillon, le modèle réflectif est meilleur que le modèle formatif à partir d'un seuil des liens structuraux. Ce résultat confirme ce qui pouvait être intuitivement considéré : plus les liens structuraux sont élevés, plus le modèle réflectif semble opérationnel comme modèle de mesure. Le choix entre les deux modèles dépend du SMC et de la taille de l'échantillon.

Ces résultats confirment que le modèle réflectif est plus efficient que le modèle formatif avec des liens structuraux élevés. Ces résultats confirment Cadogan & Lee (2013); Howell (2014); Rhemtulla et al. (2015) dans le sens où le modèle de mesure formatif peut être considéré comme un modèle de mesure réflectif avec des liens structuraux élevés. Mais dès que ces liens faiblissent, le modèle formatif garde toute sa légitimité.

4.5.3.2 Résultats sur le test LR

Ici nous analysons l'information statistique apportée par les indicateurs formatifs par un test LR.

Avec un seuil de 95% (95% des simulations valident H_0 , c'est-à-dire que le modèle formatif n'apporte pas d'information), il n'y a que très peu de configurations qui émergent. Et les résultats sont complexes à analyser car aucune structure logique sur les paramètres ne semble explicative de ces résultats. J'ai augmenté le nombre de simulations pour en faire 1000 par configuration, mais les résultats sont restés identiques.

J'ai donc pris un seuil d'acceptabilité plus lâche, 90% : 90% des répliques ne rejettent pas l'hypothèse H_0). Avec ce seuil, les résultats sont nettement plus clairs.

TABLEAU 17 : MODELE 2, MODELE 1, LIMITE INFERIEURE DES λ POUR VALIDER LE TEST LR AU SEUIL DE 0.90

$\lambda \geq$	SMC = .4	SMC = .5	SMC = .6	SMC = .7	SMC = .8	SMC = .9
N = 250	$\lambda \geq .4$	$\lambda \geq .4$	$\lambda \geq .4$	$\lambda \geq .4$	$\lambda \geq .4$	$\lambda \geq .4$
N = 500	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .3$
N = 1000	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .3$	$\lambda \geq .2$	$\lambda \geq .2$

Les indicateurs formatifs apparaissent superflus quand les liens structuraux sont supérieurs à 0.4 pour un échantillon de 250 et 0.3 pour un échantillon de 500 ou plus. Mais ceci n'est vrai qu'au seuil de 90%, ce qui veut dire que pour un nombre non négligeable de modèles les indicateurs formatifs ne sont pas superflus au regard de ce test.

En conclusion, ce test LR ne confirme que partiellement le caractère superflu des indicateurs formatifs pour le modèle de mesure.

4.5.4 Conclusion

Tous ces résultats nuancent l'hypothèse que le modèle de mesure formatif soit superflu et que le modèle opérationnel soit réflexif. Si les liens structuraux sont importants, le modèle de mesure semble indubitablement réflexif, mais dès que les liens structuraux diminuent, le modèle de mesure formatif semble plus opérant.

Ces résultats sont corroborés par l'analyse des corrélations et par le test LR, mais qui ne donne que des tendances. Donc nos résultats ne confirment que partiellement ce que considèrent Howell (2014); Lee et al. (2013); Rhemtulla et al. (2015).

4.6 Détection des mauvaises spécifications

La fiabilité d'un modèle de mesure formatif dépend de sa bonne spécification. Je ne regarde que deux problèmes centraux : les règles de proportionnalité et l'exhaustivité des indicateurs formatifs. L'objectif est de vérifier si des indices empiriques (MI et SMC) permettent de détecter un modèle de mesure incorrectement spécifié. Je ne donne que des résultats synthétiques car ces simulations sont faites sur un modèle MIMIC simple et sur un échantillon de 500. Ces résultats demanderaient à être renforcés par de nouvelles simulations pour confirmer mes premières conclusions.

4.6.1 Détection du non respect des contraintes de proportionnalité

Je rappelle que la latente formative doit médier l'intégralité de l'information des indicateurs formatifs sur les variables endogènes du modèle. Le modèle théorique est un modèle mal spécifié (Figure 34). Le modèle estimé est le modèle Figure 31. L'objectif est de vérifier si les MI détectent ce lien réel entre X1 et Y1 (intégré dans la genèse des données) qui n'est pas à priori intégré au modèle estimé.

4.6.1.1 Simulations

Ici N sera toujours à 500. Nous prenons un SMC de de 0.6 à 0.9. Ce modèle est un modèle MIMIC qui théoriquement stipule un lien entre X1 (indicateur formatif) et Y1 (variable manifeste réflexive). Les paramètres λ_1 et λ_2 représentent les corrélations entre η et, respectivement, Y1 et Y2. Dans le modèle théorique (Figure 34), c'est la corrélation globale (englobant conjointement η et X1) avec Y1 qui varie de 0.2 à 0.8 ; λ_2 varie de 0.2 à 0.8.

4.6.1.2 Résultats

Nous regardons par une statistique de comptage si au moins 95% des modèles statistiques estimés exhibent un MI supérieur au seuil de 7.88.

Les simulations montrent que dès que le lien entre X1 et Y1 *ad hoc* de la latente est élevé (α supérieur à 30%, équation (26)), la proportion des simulations indiquant un MI supérieur à 7.88 est supérieur à 95% (pour une configuration au moins 95% des simulations révèlent un MI supérieur à 7.88). Pour $\alpha = 0.3$, le taux est au minimum de 88%. Avec $\alpha < 0.3$ le risque de ne pas détecter un lien direct entre X1 et Y1 est réel.

Par conséquent lorsque la latente médiatise au moins de 30% de l'effet total des indicateurs formatifs sur les autres variables endogènes, les MI permettent de détecter une mauvaise spécification du modèle de mesure. En conclusion, pour les situations où un lien *ad hoc* entre un indicateur formatif et une variable réflexive endogène n'est pas faible, un praticien devrait pouvoir le détecter. Ce résultat est important car ce critère semble être un cadre usuel de mauvaise spécification, voir les exemples de Howell et al. (2007b); G. Kim et al. (2010); Wilcox et al. (2008) dont j'ai discuté qui souffrent de problème qui expliquerait en partie les instabilités trouvées.

4.6.2 Détection de la non exhaustivité des indicateurs formatifs

L'objectif ici est de vérifier si le SMC permet de détecter la non-exhaustivité des indicateurs formatifs.

4.6.2.1 simulations

De nouveau c'est un modèle MIMIC avec une taille d'échantillon de 500 qui est posé. Nous générons des modèles avec un SMC de la variable latente η variant de 0.1 à 0.5, qui représente une absence d'information sur la latente (il manque des indicateurs) et donc indique qu'il n'y a pas exhaustivité des indicateurs dans le modèle. L'objectif ici est de vérifier si le \widehat{SMC} permet de révéler cela.

4.6.2.2 Résultats

Concernant, la violation de l'hypothèse de l'exhaustivité des mesures formatives, les résultats montrent que la puissance statistique de l'indice d'ajustement RMSEA est relativement faible. Les résultats de nos simulations montrent que les RMSEA sont en moyenne toujours inférieurs à 0.08 (bon fit) même dans le cas de liens structurels λ faibles. Cet indicateur (RMSEA) statistique ne permet donc pas de déceler un modèle de mesure formatif incorrectement mal spécifié.

Les résultats montrent que dès que le SMC théorique est supérieur à 0.3 et les λ d'au moins de 0.4, le biais devient faible (inférieur à 5% en valeur relative). Cependant, quel que soit le cadre paramétrique, si le biais en valeur relative ($B(\widehat{SMC}) = \frac{\widehat{SMC} - SMC}{SMC}$) peut être important, ce biais en valeur absolue du SMC ($|\widehat{SMC} - SMC|$) ne dépasse jamais 0.10 (situation où le SMC théorique vaut 0.10 ; SMC estimé moyen à 0.20). Concrètement, un praticien ayant un SMC

estimé inférieur à 0.45 (seuil minimal de 0.5 usuellement accepté, enlevé du biais potentiel d'estimation) pourra considérer que son modèle de mesure est incorrectement spécifié (quand bien même il aurait un RMSEA inférieur au seuil d'acceptabilité usuel) ; à l'inverse, dans toutes les configurations, un praticien ayant un SMC au moins égal à 0.55 (seuil usuel plus le biais potentiel) sera serein pour considérer que son SMC théorique est supérieur à 0.5.

En conclusion, un praticien pourra se fier au \widehat{SMC} pour valider la bonne spécification de son modèle. Rappelons que les résultats antérieurs ont imposé un SMC supérieur à 0.6, donc un \widehat{SMC} aussi supérieur à 0.6 puisqu'à ce niveau de SMC le biais est très faible (et un \widehat{SMC} supérieur à 0.6 indique un vrai SMC supérieur à 0.6).

4.6.3 Conclusion

Ces simulations ont testé deux menaces essentielles pour un modèle formatif : 1/ le non-respect des règles de proportionnalité indiquant que la latente formative ne médiate pas l'ensemble de l'information de ses indicateurs formatifs ; 2/ l'exhaustivité des indicateurs formatifs. Pour ces deux menaces des analyses plus approfondies seraient nécessaires pour consolider nos premiers résultats. Cependant, les résultats nous autorisent à considérer qu'empiriquement, grâce aux indicateurs d'ajustement fournis par CB-SEM (les MI et le SMC), un praticien peut vérifier la bonne spécification d'un modèle de mesure formatif estimé.

4.7 PLS-SEM

PLS-SEM est souvent présenté comme l'alternative à CB-SEM pour gérer les modèles avec de variables latentes formatives. Je ne développerai pas en détail tous les résultats car l'approche PLS-SEM est largement discutée depuis quelques années, et ces simulations doivent s'intégrer dans toute la discussion à ce sujet. Or cette thèse n'a pas pour objet d'étudier l'efficacité de PLS-SEM car un tel travail exigerait une revue de littérature contradictoire actualisée... très volumineuse et qui demanderait un travail de thèse spécifique. L'objet de ce travail est juste de donner quelques éléments de repères sur PLS-SEM comparativement à CB-SEM, et donc en considérant ces résultats comme indicatifs, mais devant être consolidés.

4.7.1 PLS-SEM

La représentation PLS-SEM apparaît parfois comme une alternative pour contourner les problèmes liés à CB-SEM. Dans le cadre de la modélisation PLS-SEM, le construit est déterminé entièrement par ses indicateurs de mesures et en conséquence aucune autre cause n'impacte la variance du construit (Bollen, 2011; Cenfetelli & Bassellier, 2009; Tenenhaus et al., 2005). Ainsi, dans l'approche PLS-SEM le construit est défini comme une variable composite (VC) et n'incorpore pas l'erreur de mesure.

$$C = W_1 X_1 + W_2 X_2 + W_3 X_3 + W_4 X_4 \quad (38)$$

Pour Bollen (2011), « *At least as important, but less obvious, is that composite indicators need not share unidimensional conceptual unity [...] if we have no unidimensional theoretical concept in mind, how can we ask whether the composite indicators measure it? With composite indicators and a composite variable, the goal might not be to measure a scientific concept, but more to have a summary of the effects of several variables [...] if these are causal indicators intended to tap a single latent variable, then the coefficients of the causal indicators should be stable across different outcomes. This could explain some of the differences expressed in Howell et al. (2007) versus Bollen (2007) where the former is referring to composite (formative) indicators and the latter refers to causal indicators.* » (p. 366). Cette discussion renvoie donc à la faiblesse potentielle de PLS-SEM pour gérer les construits formatifs.

Concrètement, avec PLS-SEM, nous considérerons avoir une estimation exacte de la latente mesurée formativement ($SMC=1$), alors qu'avec CB-SEM nous aurons une estimation de la variance de l'erreur de mesure. Certes le cadre non paramétrique de modélisation PLS-SEM est moins contraignant mais il n'est pas moins sensible à l'omission de mesures importantes, à la colinéarité et à l'interprétation confondante que CB-SEM (G. Kim et al., 2010; Tensaout et al., 2011). Ce constat n'est pas une surprise car l'algorithme de PLS-SEM s'adapte aux données pour optimiser les liens linéaires entre toutes les variables du modèle, et ceci en faisant évoluer tour à tour le modèle externe et le modèle interne. Ainsi, cette approche est fortement dépendante des données observées (« *Data oriented* », Fornell & Cha, 1994). De ce fait, nous pouvons faire l'hypothèse que les paramètres du modèle de mesure formatif seront très fortement dépendants des autres variables retenues dans le SEM.

4.7.2 Simulations

PLS-SEM a été analysé dans la phase 1 des simulations.

Les paramètres de la phase 1 sont :

- N (Taille de l'échantillon) : 200, 300, 500, 750, 1000
- Le SMC variera de 0.6 à 0.9 par pas de 0.1
- Les coefficients structureaux λ sont tous identiques et varieront de 0.2 à 0.8 par pas de 0.2 ;
- R : La colinéarité entre les indicateurs formatifs varie de 0 à 0.8 par pas de 0.2.

Soit pour cette phase 1 de simulations nous avons 400 configurations (= $5 \times 4 \times 4 \times 5$) ; 5 tailles d'échantillon, 4 paramètres pour le SMC, 4 paramètres pour les λ , 5 paramètres pour R. Chaque configuration est simulée 500 fois, soit 200 000 modèles.

Dans ce qui suit, je reprends les mêmes analyses effectuées sur CB-SEM afin de comparer les performances des modélisations PLS-SEM à CB-SEM.

4.7.3 Comparaisons PLS-SEM et CB-SEM pour le critère des indicateurs formatifs

Concernant les analyses avec le « *critère des γ* » dans la phase 1 (je rappelle que ce critère était basé sur les intervalles de confiance pour Cb-SEM qui peuvent biaiser les résultats même si la normalisation était celle proposée par Franke et al., 2008). Cependant les résultats peuvent apparaître comme une indication car les tests ont été faits sur les mêmes données.

Avec la méthode CB-SEM nous avons obtenu 36,74% des simulations qui remplissent ce critère contre 43.32% avec PLS-SEM. La différence est donc significative (*paired Student test*, $p\text{-value} < 0.0001$). Concernant les résultats de l'ANOVA, là aussi ils diffèrent selon les données utilisées. Avec les données obtenues avec PLS-SEM, le seul facteur qui impacte le nombre de simulations valides est le lien structurel λ . En reprenant le même seuil de 95% que lors de notre analyse avec CB-SEM (pour un cadre paramétrique donné, plus de 95% des 500 doivent satisfaire le « critère des γ ») : PLS-SEM, contrairement à CB-SEM, conclut à l'absence d'interprétation confondante dans pratiquement toutes les configurations quels que soit N, SMC et R, et si $\lambda \geq 0.6$. Ce résultat montre que PLS-SEM est nettement moins sujet à l'interprétation confondante que CB-SEM selon le « critère des γ ».

De plus PLS-SEM, de par son mode d'estimation, n'est pas sensible à la colinéarité des indicateurs formatifs.

TABLEAU 18 : LA VALEUR λ MINIMALE NECESSAIRE A L'ABSENCE DE L'INTERPRETATION CONFONDANTE AVEC LE « CRITERE DES γ »

	R=0.0	R=0.2	R=0.4	R=0.6	R=0.8
N=200	$\lambda \geq 0.8$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$
N=300	$\lambda \geq 0.8$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$
N=500	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$
N=750	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$
N=1000	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$	$\lambda \geq 0.6$

4.7.4 Comparaison CB-SEM et PLS-SEM sur le score estimé

Si nous comparons CB-SEM et PLS-SEM sur le « critère d'identité » qui nous semble plus pertinent que le critère sur les indicateurs formatifs (corrélacion entre le score estimé de la variable latente η et le score théorique), la différence est significative mais ici en faveur de CB-SEM : la moyenne des corrélacions pour CB-SEM est de 0.8731 ; la moyenne des corrélacions pour PLS-SEM est 0.8468 (*paired Student test*, $p\text{-value} < 0.0001$). Ce résultat est fait en n'intégrant que les modèles ayant convergé sous CB-SEM. Or il n'est pas rare que CB-SEM ne converge pas avec des échantillons faibles.

Le rôle des différents paramètres influençant la qualité d'ajustement du modèle avec PLS-SEM pour ce « *critère d'identité* » sont très différents que discutés précédemment avec « *le critère des γ* » : le facteur le plus important est maintenant le SMC, le rôle de λ semble moins important qu'auparavant.

TABLEAU 19 : RESULTAT DES ANOVA AVEC PLS-SEM, A/ AVEC LE « *CRITERE DES γ* » ; B/ AVEC LES CORRELATIONS DES SCORES

	Critère des γ Rsquare=0.8226		Critère d'identité Rsquare=0.9443	
Variable	Coefficient (standardisé)	Significativité	Coefficient (standardisé)	Significativité
N	- 0.009	0.6541	0.124	<0.0001
R	0.034	0.1068	0.002	0.864
SMC	0.041	0.0508	0.927	<0.0001
λ	0.905	<0.0001	0.264	<0.0001

Si nous retenons le même seuil de 0,05 qu'avec CB-SEM, il n'y a par contre aucun modèle estimé avec PLS-SEM qui puisse être considéré comme acceptable. Avec PLS-SEM, la corrélation, entre le score estimé de la latente et celui de la latente théorique, est toujours inférieure en moyenne à 0.95 ; c'est-à-dire que PLS-SEM n'est pas meilleur sur ce critère que CB-SEM

Si nous examinons enfin « *le critère de proximité* » (la corrélation entre le score estimé et le score théorique est supérieur à 0.8), dès que le $SMC \geq 0.7$, PLS-SEM donne des estimations du score de la latente formative proche du score réel, et ceci quel que soit N. Par contre, dès que le SMC décroît ($SMC=0.6$), PLS-SEM comme CB-SEM semble peu performant. Le Tableau 20 récapitule, selon N et le SMC, les λ minimaux pour avoir au moins 95% des simulations ayant une corrélation (entre le score estimé et le score théorique) supérieure à 0.8.

TABLEAU 20 : LA VALEUR λ MINIMAL NECESSAIRE A L'ABSENCE DE L'INTERPRETATION CONFONDANTE AVEC LE « CRITERE DE PROXIMITE »

$\lambda \geq$	SMC=0.6	SMC=0.7	SMC=0.8	SMC=0.9
N=200	Aucun	0.6	0.4	0.4
N=300	Aucun	0.6	0.4	0.4
N=500	Aucun	0.4	0.4	0.2
N=750	Aucun	0.4	0.2	0.2
N=1000	Aucun	0.4	0.2	0.2

En conclusion, Nos résultats empiriques semblent montrer que PLS-SEM serait moins sensible à l'interprétation confondante que CB-SEM, ce qui confirme les résultats obtenus par Ringle et al. (2008) sur la supériorité de PLS-SEM vs CB-SEM pour estimer les modèles de mesure formatifs.

4.7.5 Comparaison de CB-SEM et PLS-SEM sur le modèle structurel

Nous examinons maintenant le « critère des λ », en moyenne, CB-SEM et PLS-SEM ne sont pas différents: Biais moyen avec CB-SEM=0.1305 et Biais moyen avec PLS-SEM=0.1289 (test de Student apparié : p-value=0.79). Cependant, si nous reprenons le seuil de 5% d'erreur de biais (Hoogland & Boomsma, 1998), CB-SEM est préférable à PLS-SEM. En effet, avec PLS-SEM, il faut un SMC = 0.9 et un $\lambda \geq 0.8$ pour que le biais moyen relatif soit inférieur à 5%. De plus, pour PLS-SEM, le biais absolu ($|\hat{\lambda} - \lambda|$) pour chaque λ peut être de 0.2, soit une source de confusion potentielle pour le praticien. Le Tableau 21 montre un comparatif des liens $\hat{\lambda}$ minimaux nécessaires entre la latente formative et les variables réflectives endogènes pour avoir un biais relatif moyen inférieur à 5%.

TABLEAU 21 : LA VALEUR MINIMALE DE λ MINIMAL POUR OBTENIR UN BIAIS RELATIF D'ESTIMATION λ INFÉRIEUR A 5%, EN FONCTION DE LA TAILLE DE L'ÉCHANTILLON ET DU SMC

$\lambda \geq$	Modèle	SMC=0.6	SMC=0.7	SMC=0.8	SMC=0.9
N=200	CB-SEM	0.6	0.6	0.6	0.6
	PLS-SEM	Aucun	Aucun	Aucun	Aucun
N=300	CB-SEM	0.6	0.6	0.6	0.6
	PLS-SEM	Aucun	Aucun	Aucun	Aucun
N=500	CB-SEM	0.4	0.4	0.6	0.4
	PLS-SEM	Aucun	Aucun	Aucun	Aucun
N=750	CB-SEM	0.4	0.4	0.4	0.4
	PLS-SEM	Aucun	Aucun	Aucun	0.6
N=1000	CB-SEM	0.4	0.4	0.4	0.4
	PLS-SEM	Aucun	Aucun	Aucun	0.6

CB-SEM est plus performant que PLS-SEM, dès lors que le modèle est bien spécifié, quant à l'estimation du modèle structurel.

4.7.6 Conclusion

En conclusion, PLS-SEM paraît préférable selon le « *critère des γ* ». De même, un modèle estimé avec PLS-SEM semble plus performant avec le « *critère de proximité* ». Nos résultats empiriques montrent que PLS-SEM serait moins sensible à l'interprétation confondante du modèle de mesure que CB-SEM, ce qui confirme les résultats obtenus par Ringle et al. (2008) sur la supériorité de PLS-SEM vs CB-SEM pour estimer les modèles de mesure formatifs. Cependant ces résultats sont à nuancer car le modèle structurel (les coefficients des liens structuraux) risque d'être fortement biaisé avec PLS-SEM.

De plus notons qu'un praticien ne peut s'assurer avec PLS-SEM de la fiabilité des estimations des λ , en effet il faut avoir un SMC =0.9 et $\lambda =0.8$, or PLS-SEM ne fournit pas le SMC (rappelons que la méthode PLS-SEM ignore l'erreur associée à la latente formative). Donc un praticien n'a aucun moyen de vérifier empiriquement s'il est dans une configuration ($\widehat{SMC} \geq 0.9$) lui permettant de considérer que les $\hat{\lambda}$ (qui doivent alors être égaux à 0.8 au moins) ne sont pas biaisés, seul moyen de s'assurer empiriquement de la stabilité des

paramètres γ et λ . De plus un praticien n'a aucun outil avec PLS-SEM pour vérifier la bonne spécification d'un modèle estimé par PLS-SEM (le respect des règles de proportionnalité et l'exhaustivité des indicateurs formatifs). La littérature académique soulève en conséquence un certain nombre de limites à PLS-SEM relativement à l'absence de l'erreur associée à la variable latente formative (Diamantopoulos, 2011; Marcoulides et al., 2009). Concrètement, un praticien ne peut utiliser les indicateurs statistiques utilisés pour la discussion de nos hypothèses avec PLS-SEM. Cette limite handicape fortement PLS-SEM.

En effet, comme l'affirme Bollen (2007) et comme le montrent nos simulations, l'interprétation confondante se révèle être en grande partie la conséquence d'un problème de spécification du modèle estimé. PLS-SEM est alors moins préférable à CB-SEM car PLS-SEM ne produit pas d'indice d'ajustement permettant de tester la qualité d'ajustement du modèle. Or, le praticien a justement besoin d'indicateurs d'ajustement lors de la phase de validation (Chen et al., 2008; Fan et al., 1999; Steiger, 2007).

En conclusion, si les paramètres du modèle de mesure apparaissent dans certaines configurations moins biaisés avec PLS-SEM qu'avec CB-SEM, ceux du modèle structurel sont fortement biaisés, à l'inverse de la méthode CB-SEM. Nous retrouvons un résultat usuel dans les comparaisons entre CB-SEM et PLS-SEM. CB-SEM est supérieur pour estimer le modèle structurel, PLS-SEM est préférable pour estimer le modèle de mesure.

4.8 Discussion générale

Je synthétise les résultats de l'ensemble des simulations pour donner un cadre opérationnel de l'usage d'un modèle de mesure formatif... toujours dans l'hypothèse empiriste-réaliste qui considérerait comme une réalité une propriété mentale sans manifestation et ne pouvant être inférée que par des indicateurs formatifs.

4.8.1 Les résultats de nos simulations

Le modèle de mesure formatif montre, qu'avec un certain niveau de liens structuraux, les indicateurs formatifs ne sont pas « instables » (si ces derniers ne soient pas colinéaires). De même le score estimé est fiable en termes de « proximité » du vrai score, mais pas identique au vrai score (corrélation entre 0.80 et 0.95 en général). Ceci est vrai avec CB-SEM et PLS-SEM. La fiabilité du modèle de mesure dépend du nombre de variables réflectives endogènes, plus ce nombre est grand plus le modèle de mesure formatif sera fiable.

Ces résultats nuancent donc les propos des opposants aux modèles de mesure formatifs dans le sens où ces derniers n'exhibent pas une instabilité « congénitale ». Néanmoins ces résultats confirment la dépendance du modèle de mesure formatif aux variables endogènes, et ceci avec CB-SEM et PLS-SEM.

En conséquence, l'hypothèse de considérer le modèle de mesure formatif comme un modèle de mesure réflectif (de second ordre) méritait d'être testé par simulations. De nouveau nos résultats sont nuancés. A un certain niveau de liens structuraux, le modèle de mesure semble être plus efficient si on le considère réflectivement et en conséquence les indicateurs formatifs sont en effet superflus. Mais avec des liens structuraux plus faibles les indicateurs formatifs apportent de l'information et on ne peut considérer le modèle de mesure comme réflectif.

De plus, un modèle réflectif de second ordre ne peut être identifié avec seulement deux latentes réflectives de second ordre, donc les indicateurs formatifs sont nécessaire pour un tel modèle pour son identification.

4.8.2 Le cadre d'utilisation d'un modèle de mesure formatif

Le bilan de toutes ces simulations montre que la fenêtre paramétrique est faible pour avoir : 1/ un modèle de mesure formatif fiable ; 2/ un modèle de mesure qui ne puisse être considéré comme un modèle réflectif.

Le tableau ci-dessous rapporte ce cadre paramétrique

TABLEAU 22: GUIDE POUR DEVELOPPER UN MODELE DE MESURE EFFICIENT AVEC CB-SEM

Model	Modèle de mesure le plus efficient	N	$0.4 \leq SMC < .7$	$SMC = .7$	$SMC = .8$	$SMC = .9$
2 latentes réflectives endogènes	MMF	250	/	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .9$
		500		$\lambda \geq .6$	$\lambda \geq .7$	$\lambda \geq .8$
		1000		$\lambda \geq .5$	$\lambda \geq .6$	$\lambda \geq .7$
	MMR	250-500-1000	/	/	/	/
4 latentes réflectives endogènes	MMF	250	/	$0.4 \leq \lambda < .7$	$0.3 \leq \lambda < .8$	$0.2 \leq \lambda < .9$
		500		$0.3 \leq \lambda < .7$	$0.2 \leq \lambda < .8$	$0.2 \leq \lambda < .9$
		1000		$0.2 \leq \lambda < .7$	$0.2 \leq \lambda < .8$	$0.2 \leq \lambda < .9$
	MMR	250	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .8$	$\lambda \geq .9$
		500	$\lambda \geq .7$	$\lambda \geq .7$	$\lambda \geq .8$	$\lambda \geq .9$
		1000	$\lambda \geq .6$	$\lambda \geq .7$	$\lambda \geq .8$	$\lambda \geq .9$

MMF : Modèle de mesure formatif ; MMR : modèle de mesure réflectif

J'ai discuté des approches CB-SEM et PLS-SEM, les deux ayant leurs avantages et limites. Je donne pour le modèle 1 (avec deux latentes réflectives) le cadre opérationnel de telles mesures.

TABLEAU 23 : TABLEAU RECAPITULATIF DES COMPARAISONS ENTRE LES METHODES CB-SEM ET PLS-SEM

	CB-SEM	PLS-SEM
Vérification du rapport de proportionnalité entre les γ et les λ	Test valide (MI)	Pas de test
Vérification de l'exhaustivité des indicateurs formatifs	Test valide (SMC)	Pas de test
Cadre pour éviter une interprétation confondante au niveau de l'estimation des γ	$N \geq 300$; $R \leq 0.6$; $\lambda \geq 0.6$.	$\lambda \geq 0.6$
Cadre pour éviter une interprétation confondante au niveau de la mesure de η	1/ Pour avoir une identité entre la latente estimé et la latente théorique :	1/ Pour avoir une identité entre la latente estimée et la latente théorique : aucune

	<p>a/ pour $N < 750$: $SMC = 0.9$ et $\lambda = 0.8$;</p> <p>b/ pour $N \geq 750$: $\lambda \geq 0.6$.</p> <p>2/ Pour avoir une estimation de la latente formative proche de la latente théorique : $\lambda \geq 0.6/0.8$ selon les configurations (cf. Tableau 11 ; Tableau 14)</p>	<p>configuration</p> <p>2/ Pour avoir une estimation de la latente formative proche de la latente théorique : $SMC \geq 0.7$ et $\lambda \geq 0.2/0.4/0.6$ selon les configurations (cf. Tableau 20 ; Tableau 21)</p>
Cadre pour éviter une interprétation confondante au niveau des estimations des λ	$\lambda \geq 0.4/0.6$ selon les configurations (cf. Tableau 13)	<p>$SMC \geq 0.9$</p> <p>$\lambda \geq 0.8$</p>

Pour estimer un tel SEM avec une variable latente formative exogène, PLS-SEM semble plus performant que CB-SEM pour estimer les paramètres γ , mais PLS-SEM possède plusieurs faiblesses 1/ PLS-SEM produit des estimations des λ nettement plus biaisées que CB-SEM, et 2/ PLS-SEM ne permet pas de juger de la qualité d'ajustement du modèle de mesure et plus généralement du SEM. De plus, ce qui doit être au cœur de la discussion dans un MES, c'est le modèle structurel, c'est-à-dire les liens entre la latente formative η et les autres latentes endogènes (Figure 32). En ce sens CB-SEM est préférable à PLS-SEM pour un praticien. Cependant, une solution pratique pourrait être de fonctionner en plusieurs étapes : 1/ Estimer un SEM avec CB-SEM ; 2/ Juger la qualité d'ajustement du modèle de mesure et les biais potentiels des λ à partir de indicateurs de CB-SEM ; 3/ lorsque le SEM est jugé acceptable ; 4/ Réestimer le SEM avec PLS-SEM pour minimiser les biais d'estimation des γ .

4.8.3 L'exemple de Kim et al. 2011

À la lumière de nos résultats, nous pouvons réexaminer ceux de G. Kim et al. (2010). Ces auteurs ont rapporté différents exemples empiriques pour montrer que la mesure formative est sujette à interprétation confondante. Je reprends donc cet exemple pour caractériser en quoi ce n'est pas de façon générique le modèle formatif qui est instable, mais le besoin d'user d'un cadre paramétrique précis que ne respecte pas ces exemples.

Nous discutons du seul cas 1, modèle F 1 et F-modèle 2 (G. Kim et al., 2010, p. 351). L'étude a inclus 243 réponses, ce qui est un petit échantillon pour un modèle formatif. Les seuils pour obtenir des résultats fiables (sur la base de nos simulations de la phase 1 qui intégraient la colinéarité des indicateurs formatifs) sont $\lambda \geq 0,8$; $SMC \geq 0,6$; $R \leq 0,2$. La matrice de corrélation des indicateurs formatifs (G. Kim et al., p. 351) donne plusieurs valeurs supérieures à 0,2; ce qui rend le modèle instable. Avec CB-SEM, les estimations de λ standardisées des exemples de ces auteurs sont inférieures à 0,6 et parfois proche de zéro, qui peut également caractériser un modèle susceptible d'être instable. Les SMC sont respectivement de 0,34 et 0,35 (voir le Tableau 6) qui caractérise un modèle mal spécifié car l'exhaustivité des indicateurs formatifs n'est pas validée. Comme nous l'avons montré (Tableau 5), il y a $MI > 7.88,84$, de sorte que nous ne pouvons pas postuler que la contrainte de proportionnalité soit respectée, et donc la variable latente η ne médiate pas complètement les effets des indicateurs formatifs sur les variables latentes endogènes, ce qui caractérise un modèle mal spécifié. Tous ces résultats permettent de comprendre pourquoi les modèles présentés par Kim et al. sont certainement instables.

5 Conclusions de la partie 2

Plusieurs experts recommandent d'éviter les modèles de mesure formatif (Bagozzi, 2011b; Edwards, 2011; Hardin & Marcoulides, 2011; Howell, 2014; Lee et al., 2013; Markus, 2014). Dans des études antérieures, la question de l'interprétation confondante a été examinée par la dépendance de la variable latente formative aux latentes réflectives endogènes conduisant à l'instabilité des paramètres formatifs (Franke et al., 2008; Howell, 2014; Howell et al., 2007b; G. Kim et al., 2010; Wilcox et al., 2008). Mais comme je l'ai discuté, ces exemples étaient source de confusion autant sur les données employées que sur les critères retenus. Par ailleurs, des arguments théoriques prétendent que le sens de la variable latente avec les mesures formatives est donné par le modèle réflectif, ce qui signifie que les indicateurs formatifs sont superflus pour le modèle de mesure (Cadogan & Lee, 2013; Howell, 2014; Lee et al., 2013; Rhemtulla et al., 2015; Wilcox et al., 2008). Mais aucune simulation n'avait été entreprise pour vérifier cette hypothèse.

Tout mon travail apporte des réponses clarifiées sur l'ensemble de ces problèmes.

5.1 Modèles formatifs à éviter

Les résultats de mon travail montrent que les modèles de mesure formatifs ne sont pas intrinsèquement soumis à interprétation confondante si le modèle de mesure est correctement spécifié (une exhaustivité des indicateurs formatifs et des liens structuraux importants) et si les conditions de taille de l'échantillon peuvent être satisfaites. De même, les simulations montrent que le modèle de mesure fiable n'est pas forcément le modèle réflectif. Il y a un petit spectre de paramètres pour lesquels un modèle de mesure formatif est stable, fiable et pour lesquels le modèle de mesure réflectif donne des scores estimés moins efficaces que le modèle de mesure formatif. Ce résultat important conteste ce que des auteurs considèrent comme l'instabilité intrinsèque des modèles de mesure formatifs, mais a contrario confirme leur propos concernant l'importance des variables endogènes.

Cependant, en ligne avec Diamantopoulos (2011), il y a de grandes difficultés à développer un modèle de mesure formatif valide à partir des impératifs inhérents, à savoir : l'exhaustivité des indicateurs formatifs, faible colinéarité entre eux, le respect des contraintes de proportionnalité, des liens structuraux suffisamment importants mais sans faire glisser le

modèle de mesure vers un modèle réflectif. A tous ces problèmes s'ajoute des hypothèses non-testées comme l'obtention d'indicateurs formatifs sans aucune erreur de mesure.

In fine, même du point de vue empiriste-réaliste, les modèles de mesure formatifs semblent réservés à des situations assez rares. Plutôt que de construire un modèle de mesure qui aura toutes les chances de se révéler peu fiable, il me paraît nettement préférable de partir d'un modèle de mesure réflectif même dans le cadre du réalisme-empirique.

Pour les situations où l'on ne peut pas mesurer réflectivement une propriété mentale... la discussion renvoie à ce que j'ai développé en introduction : que mesure-t-on alors avec le modèle formatif ? Car la propriété mesurée n'a pour seule fonction que d'impacter d'autres propriétés non directement observables. Il y a donc en définitive autant sur le plan ontologique que méthodologique bien des raisons de considérer que les modèles de mesure formatifs sont à proscrire.

5.2 CB-SEM ou PLS-SEM

Dans les très rares situations où un modèle formatif serait impératif, et sous hypothèse d'être dans le spectre de paramètres permettant une fiabilité des résultats, la question est de savoir s'il faut user de CB-SEM ou de PLS-SEM.

L'approche de CB-SEM semble perdre du terrain en faveur de l'approche PLS-SEM pour modéliser les construits formatifs. Mais c'est sous-estimer les avantages de CB-SEM, particulièrement pour comparer des modèles nichés (Bollen, 2011; Diamantopoulos, 2011). Or, le processus de validation du modèle formatif nécessite de tester plusieurs hypothèses rivales afin de minimiser les risques d'une mauvaise spécification. Cette flexibilité à examiner différentes hypothèses théoriques n'a pas d'équivalent dans l'approche PLS-SEM. PLS-SEM étant non-paramétrique, il peut apporter des réponses dans des situations où CB-SEM n'est pas fonctionnelle. Le cadre non paramétrique de PLS-SEM, qui est un avantage par rapport à CB-SEM, fragilise *a contrario* l'usage de modèles formatifs, car la stabilité de ses coefficients dépend de la bonne spécificité du modèle global, information que PLS-SEM ne peut pas fournir au chercheur. Même si les tests d'ajustement fournis par CB-SEM ont une puissance statistique faible, ils constituent néanmoins des indicateurs précieux lors de la phase de validation (Chen et al., 2008; Fan et al., 1999; Steiger, 2007). De plus, comme discuté précédemment, le statut d'une variable composite (formative) avec PLS-SEM reste

conceptuellement très problématique et est la raison essentielle de son rejet comme solution alternative à CB-SEM pour différents auteurs.

Par ailleurs, l'algorithme de PLS-SEM peut faire penser que PLS-SEM serait moins sensible aux violations des hypothèses de CB-SEM sur les distributions d'erreurs, à la taille de l'échantillon et à une mauvaise spécification du SEM. Bien que les travaux sur ces points soient rares, plusieurs simulations ont montré que PLS-SEM n'était pas moins sensible (Cenfetelli & Bassellier, 2009; Pelagatti et al., 2012; Petter et al., 2007; Ringle et al., 2009; Tensaout et al., 2011). Avec des construits formatifs il faudra avoir des indicateurs gaussiens comme avec CB-SEM, et une taille d'échantillon assez grande. Ce n'est qu'avec des construits réfléchitifs que PLS-SEM peut relâcher ces hypothèses.

Au final, PLS-SEM ne semble guère apporter des réponses efficaces aux problèmes rencontrés avec les construits formatifs avec CB-SEM. De plus, et de façon plus fondamentale, le statut d'une variable latente composite avec PLS-SEM reste conceptuellement très problématique car nous travaillons avec un *index* (simple combinaison linéaire des variables manifestes, équation (38). La question fondamentale qui se pose avec cette approche est de savoir si nous estimons une mesure formative d'un construit avec un tel *index* ? Un certain nombre d'auteurs récusent ainsi l'usage de PLS-SEM avec des mesures formatives car pour eux un *index* ne peut prétendre être une estimation d'un construit par des indicateurs formatifs à cause de l'absence du terme d'erreur de la variable latente (Diamantopoulos, 2011; Hardin & Marcoulides, 2011). Hardin et Marcoulides (2011) alertent d'ailleurs les chercheurs sur l'engouement, non justifié pour ces auteurs, pour PLS-SEM pour estimer des modèles avec des mesures formatives. Mais à l'inverse, d'autres auteurs considèrent ce débat comme erroné (Rigdon, 2012)

5.3 Recommandations

Comme je l'ai souligné, le développement d'un modèle de mesure formatif dans un SEM se heurte à des obstacles difficilement contournables. Ma première recommandation est que, même dans le cadre de l'empirisme-réaliste, chaque fois que cela est possible, il est préférable de développer des mesures réfléchitives valides. Mais il semble exister dans ce cadre réalisme-formatif des situations dans lesquelles le praticien ne possède que des indicateurs formatifs et souvent le chercheur trouve dommage de ne pas s'en servir dans les applications empiriques. Je considère que, pour de telles situations, le cadre théorique est mal posé.

Avec CB-SEM ou PLS-SEM, le praticien devra s'assurer que les indicateurs retenus de la variable latente formative sont exhaustifs (on n'oublie pas un indicateur important), qu'ils sont peu colinéaires et gaussiens. Il devra s'assurer d'avoir un échantillon important (au moins 300 personnes). Comme les indicateurs formatifs sont généralement issus de questionnaires, ils comportent par nature une erreur de mesure, donc il faudra s'assurer par une réflexion en amont que ces indicateurs n'intègrent qu'une très faible erreur potentielle. Il devra s'assurer que ces indicateurs formatifs n'expliquent pas directement les autres variables endogènes entrées dans le modèle car le construit formatif doit impérativement médiatiser l'influence de ses indicateurs formatifs sur les autres variables endogènes. Il faudra vérifier que les coefficients des indicateurs formatifs ne changent pas significativement lorsque l'on modifie les variables réflectives endogènes (en s'assurant que ces dernières sont fortement liées au construit d'intérêt). Il faudra aussi vérifier que les contraintes de proportionnalités sont respectées, car elles sont essentielles dans le processus de validation empirique (ce qui peut se faire avec les tests sur les modifications d'indices sous CB-SEM, mais pas avec PLS-SEM).

Avec CB-SEM, il est conseillé d'essayer de retirer des covariances entre les indicateurs formatifs (par des tests sur des modèles nichés). Il faut impérativement lier le construit mesuré formativement à au moins deux autres variables (ou construits) réflectifs endogènes, en s'assurant que ces dernières n'ont pas de liens indépendamment de ce construit (faisable uniquement par une réflexion théorique). Il faudra de plus ne retenir que les coefficients standardisés après analyse de leur significativité pour les praticiens n'utilisant pas *Mplus* et vérifier par des test LR la significativité des coefficients

PLS-SEM ne devra être choisi que lorsque CB-SEM n'est pas utilisable, ne serait-ce que parce qu'aucune erreur de mesure de la variable latente n'est intégrable dans PLS-SEM. En fait, les cas où PLS-SEM est préférable à CB-SEM sont rares. Si l'échantillon est faible ou les données non-gaussiennes, PLS-SEM n'est pas moins sensible que CB-SEM pour des construits formatifs. Donc PLS-SEM devra n'être utilisé que lorsque le nombre de variables intégrées est trop important pour CB-SEM, en s'assurant un échantillon important et des données gaussiennes pour les indicateurs formatifs.

Ma conclusion se veut donc une mise en garde sur le recours aux mesures formatives même dans le cadre empiriste-réaliste. Il existe un faisceau de preuves empiriques et d'arguments théoriques qui soulignent les difficultés à modéliser de ce type de mesures. Si cette discussion est amorcée sur le plan épistémologique (Bagozzi, 2011a; Borsboom et al., 2003; Markus & Borsboom, 2013b), elle semble avoir eu peu d'échos sur le terrain de l'opérationnalisation des

construits formatifs (Finn & Wang, 2014). En considérant l'ensemble des problèmes épistémologiques et méthodologiques liés aux modèles formatifs, je m'associe donc aux conclusions de Bagozzi (2007) qui indique que la relation entre des indicateurs formatifs et le construit « *n'est pas seulement conceptuellement ambiguë mais aussi empiriquement ambiguë* » (p 234), l'auteur ajoutant : « *Personnellement, je ne pense pas que le sens d'un construit mental puisse être aussi interprétable et utilisable dans une forme formative que dans une forme réflexive.* » (p 234). Ce constat s'applique aussi bien à CB-SEM qu'à l'approche PLS-SEM. Hardin & Marcoulides (2011) rappellent les divergences entre les défenseurs et les opposants à ce type de mesures. Pour Hardin et Marcoulides : « *La prudence conseillerait de suspendre temporairement l'usage de mesures formatives* » (p 1) en attendant que la recherche ait tranché sur la pertinence de leur usage. Je préconise aussi le développement des mesures réflexives lorsque cela est possible car moins sujette que les mesures formatives aux problèmes d'identification, d'erreur de mesure et d'interprétation.

Concernant l'approche PLS-SEM comme solution alternative à CB-SEM, il semble que les avantages attendus sont loin d'être atteints. Cependant des auteurs pointent les avantages de PLS-SEM pour certains modèles avec des construits formatifs. En effet comme le soulignent Ringle et al. (2012), PLS-SEM est la seule méthode utilisable lorsque nous avons des données secondaires de tailles faibles, avec parfois seulement un ou deux indicateurs disponible, et de surcroît sans la possibilité de recourir à des indicateurs réfléchifs (pas de données disponibles), le chercheur n'a plus alors le choix. Cependant, Ringle et al. (2012) préconisent lorsque cela est possible la validité d'un modèle formatif par le modèle de redondance proposé par Chin (2010), qui s'appuie sur le même construit mesuré réfléchivement. N'est-il pas plus valide alors d'utiliser le modèle de mesure réfléchif utilisé dans le modèle de redondance pour le SEM (Tensaout et al., 2011) ? Si les données pour les indicateurs réfléchifs ne sont pas disponibles, la prudence semble valoir comme le souligne Hardin & Marcoulides (2011) car PLS-SEM demande certainement d'autres travaux académiques avant de pouvoir apparaître comme une alternative fiable à CB-SEM pour de tels cas.

L'objet de cette discussion n'est pas de récuser l'usage des mesures formatives si on se place dans le cadre du réalisme-empirique. Bagozzi (2011b, p. 275) dit « *les mesures formatives et les construits formatifs ont une place dans la recherche, mais c'est crucial de reconnaître que leurs applications sont limitées à quelques modèles bien définis* ». Si donc un chercheur ne peut éviter de devoir développer un modèle de mesure formatif, il lui faudra valider toutes les hypothèses nécessaires à une telle formalisation, dont seules quelques-unes sont testables

empiriquement. C'est avant tout la théorie qui justifiera un tel modèle, avec de faibles moyens statistiques pour le valider empiriquement.

Conclusion générale

L'objet de ma thèse voulait discuter des variables latentes sur deux axes : épistémologie et méthodologie. A son amorce, je n'avais aucune idée de la trajectoire que j'allais suivre, d'où un double travail en parallèle pour essayer de donner des réponses aux deux questions centrales que me renvoyait la littérature académique sur les modèles variables latentes : 1/ quelle est la légitimité scientifique des modèles avec variables latentes ; 2/ si ces modèles sont légitimes, peut-on user de modèles de mesure formatifs.

Ce n'est qu'en fin de cheminement (et mes directeurs de thèses pourront le confirmer) que j'en suis arrivé à poser la thèse centrale que les modèles avec variables latentes ont une légitimité scientifique pour peu que l'on se réfère à un cadre épistémologique décalé par rapport au cadre dominant. Ce développement sur la légitimité des variables latentes renvoie in fine à l'illégitimité des modèles formatifs. Ma conclusion sera en conséquence essentiellement axée sur la première partie de cette thèse.

La seconde partie n'a en effet pour fonctions dans cette thèse que : 1/ illustrer la démarche vérificationniste en statistique par des simulations, montrant la difficulté de ce faire pour donner du sens opérationnalisable ; 2/ d'apporter des réponses empiriques non-publiées sur une question centrale discutée dans la littérature académique pour les auteurs récusant le cadre de pratique en psychologie que je propose (le pragmatisme-réaliste) et qui affirmeraient se référer à l'empirisme-réaliste. Même dans ce cadre épistémologique les modèles formatifs, bien qu'ayant, du point de vue réaliste-empirique, une légitimité ontologique, souffrent de biais potentiels d'estimation en dehors d'une petite fenêtre paramétrique peu en phase avec la pratique (au regard des publications usant de modèles formatifs).

Le fétichisme de la mesure en psychologie

La mesure en psychologie est le plus souvent inscrite dans un cadre empiriste-réaliste, c'est-à-dire dans une démarche scientifique qui prétend mesurer une propriété mentale indépendamment de la praxis de cette propriété (contexte, enjeux, ...). Cette démarche a le fantasme de pouvoir poser un jour une mesure (test ou une échelle de mesure) qui soit opérationnelle en tout temps et en tout lieu, prétendant mesurer un objet « réel » indépendant de sa pratique sociale. Une telle prétention se confronte de prime abord à l'impossibilité concrète de poser une conceptualisation définitive d'une propriété mentale. Toute l'histoire de la mesure en psychologie est l'histoire de révision des tests et échelle de mesure. D'un point de vue empiriste-réaliste, c'est-à-dire d'un point de vue épistémologique se référant à la

physique comme modèle, toute cette démarche est une succession d'échecs car à aucun moment la psychologie n'est capable de dire d'une façon définitive « *cette propriété mentale c'est cela et elle se mesure comme cela* ». De plus, dans le cadre empiriste-réaliste, mesurer exige de valider la mesure par un cadre formel défini par la physique ; et cela est impossible.

Ainsi, si la psychologie n'affirme pas sa rupture avec le cadre épistémologique empirique-réaliste, elle s'enferme dans une démarche de mesure qui devient effectivement du « fétichisme » comme le dit Vautier (2015c) : « *La question qui se pose aux psychologues qui calculent et interprètent des scores psychotechniques est de savoir s'ils sont prêts à assumer que certaines pratiques de fabrication et d'utilisation de scores psychotechniques soient des pratiques fétichistes. [...] Un fétiche est un objet qui permet d'évoquer, ou d'invoquer si on s'adresse à elle, une entité métaphysique par le seul biais d'une convention ; il suffit que (presque) tout le monde croie à l'entité et croie que l'objet est relié à l'entité pour que cet objet devienne un fétiche [...] Dans le domaine de la recherche, toute hypothèse dont une variable dépendante ou indépendante prend ses valeurs dans une échelle de scores vus comme des mesures n'est pas une hypothèse scientifique mais une hypothèse fétichiste. Les grandeurs psychologiques ne sont pas des concepts scientifiques tant qu'elles ne sont pas mesurables et les études de validation des tests utilisés ne testent pas la mesurabilité* ».

Je suis en accord avec Vautier sur ce fétichisme car dans le cadre empirique-réaliste, la psychologie est incapable d'inscrire sa démarche de mesure dans le cadre d'une validation scientifique en phase avec le cadre épistémologique dans lequel elle prétend inscrire sa démarche scientifique (l'empirisme-réaliste) ; donc une mesure devient « *une entité métaphysique par le seul biais d'une convention* ». Mais ce fétichisme existe car il est inscrit dans une pratique sociale de l'expertise des psychologues dans des processus de normalisation et hiérarchisation des individus. Ce n'est que par la critique de la fonction sociale actuelle de la psychométrie que l'on peut construire une autre pratique basée sur un pragmatisme-réalisme.

Pour une mesure objective des mesures subjectives

Mon point de vue défend la thèse critique de la mesure en tant que pratique inscrite dans un cadre épistémologique dominant (l'empirisme-réaliste). Pour autant, je suis aussi critique de la critique par essence de toute mesure en psychologie. En mesurant des propriétés mentales, nous ne mesurons pas des propriétés de l'être biologique (par exemple par la mesure de

processus neuronaux via la neuro-imagerie), mais des propriétés de l'être social, c'est à dire un être actif inscrit dans des interactions sociales ; des propriétés mentales inscrites dans une intersubjectivité. C'est la nature particulière des objets mesurés en psychologie qui me laisse considérer qu'il faut un décalage épistémologique et rompre avec l'empirisme-réaliste dominant, et rompre aussi avec le constructivisme. Il faut affirmer une démarche épistémologique adaptée aux objets de la psychologie. Etant des objets sociaux, ayant une fonction sociale, n'étant caractérisable que par leurs manifestations sociales, réelles mais catégorisées par la pratique sociale (ce qui les rend mouvants ne serait-ce que par l'effet de boucle tel que défini par Hacking), les propriétés mentales n'ont pas la même nature de réalité que des objets physiques extérieurs à nos relations sociales. Il faut affirmer un cadre épistémologique qui considère que les propriétés mentales sont à renégocier en permanence, qu'il n'y a pas de « vérité », juste une « objectivité pratique pour nous », c'est-à-dire une objectivité inscrite dans une pratique sociale. Ce n'est que ce décalage épistémologique qui nous permet de rompre avec le formalisme issu de la physique dans lequel s'enferment les psychologues empiriste-réaliste.

Les propriétés mentales me paraissent potentiellement mesurables pour autant que leur mesure soit inscrite dans cette renégociation. Cela veut dire concrètement inscrire une mesure non par la réplication d'un test « validé » antérieurement, mais par une démarche pratique de la construction d'un test en référence au contexte et enjeux de la mesure, affirmant mesurer une propriété en tant que propriété inter-subjective. Si pour certaines propriétés « simples » (fatigue, douleur, ...) une telle renégociation ne devrait pas bousculer le cadre d'un protocole de mesure usuel ; pour d'autres propriétés mentales (dépression, anxiété, intelligence, ...), la littérature académique offre une protéiformisation des concepts qui exige cette renégociation concrète. C'est cette renégociation permanente et nécessaire relativement à la catégorisation pratique du moment, c'est l'exigence de l'unidimensionnalité et de l'indépendance locale relativement au cadre de la théorie afférente à la mesure, qui me laissent considérer que des « échelles de mesure » peuvent plus aisément être développées que des « tests » d'habileté. Mais ce point mériterait un travail spécifique.

Un second leurre usuel dans la pratique en psychologie est l'assimilation du modèle statistique au modèle théorique. La littérature académique abonde pourtant de travaux montrant les limites des indicateurs statistiques validant un modèle statistique. Le sens des résultats ne peut s'inscrire que dans une discussion, à l'image des approches qualitatives. Affirmant que le travail de modélisation statistique est inscrit dans des choix et des heuristiques (Falissard et al., 2013). Ce n'est donc pas le modèle statistique qui valide la

mesure, mais bien une démarche pratique en amont qui permet d'objectiver une propriété comme mesurable.

De plus, comme discuté, le résultat d'une modélisation statistique génère la tension entre « résultat moyen » et « singularité ». Il n'y a pas forcément contradiction entre les deux, mais ceci doit aussi s'inscrire dans une discussion. Par contre il y a une « hérésie » usuelle qui est de considérer que l'on peut sans discussion appliquer à l'individu la dynamique moyenne qu'un modèle exhibe.

Ainsi si la mesure me paraît potentielle dans une pratique scientifique en psychologie, cette démarche doit s'inscrire dans une pratique critique.

Le pragmatisme comme méthode

J'ai usé du terme pragmatisme-réaliste, en affirmant bien de pas encadrer ma thèse dans un courant philosophique figé mais bien comme une méthode. J'aurais pu me référer au *constructivisme-réaliste* ou au *new pragmatism*, mais j'ai déjà expliqué les raisons de mon choix sémantique. Cette méthode que j'ai nommée pragmatisme-réaliste se positionne, comme le courant dominant en psychologie qui est l'empirisme-réaliste, sur la base du réalisme. Mais ce réalisme associé au pragmatisme se différencie de ce courant dominant par l'affirmation que la psychologie n'a pas à épouser le cadre de l'empirisme-réaliste tel que les sciences de la nature cadrent leur démarche.

Les objets étudiés en psychologie sont des réalités objectivées par le prisme de la subjectivité, comme tout objet étudié en science. Mais à la différence des sciences de la nature, les objets en psychologie ne peuvent en aucun cas être considérés comme des objets fixes à découvrir. Il y a une rupture avec le cadre formel des sciences de la nature dont « *la conception standard est celle d'une science conçue comme la découverte de faits qui existent "dans le monde".* » (Hacking, 2008, p. 88). Les objets étudiés en psychologie sont des objets mouvants car ils s'inscrivent comme objets d'interaction sociale. Ils ne peuvent prétendre avoir une *ontologie objective* au sens où Searle le considère ; pour reprendre les concepts de Searle, leur ontologie est uniquement subjective. Les effets de boucle (Hacking) rendent ces objets mouvants, obligeant à renégocier leurs catégories et en conséquence leurs mesures. C'est bien la particularité des objets en psychologie qui m'a fait considérer que la méthode opérante devait se référer à une épistémologie qui s'affirme réaliste tout en se démarquant du réalisme-

empirique, voyant bien la réalité comme une pratique de la connaissance, d'où ma référence au pragmatisme.

On parle de *new pragmatism* auquel on réfère Hacking, en opposition à Putnam qui serait *néo-pragmatist*. Je ne cherche pas à m'affilier à un courant de la philosophie des sciences, considérant que l'interaction des contradictions est bien plus riche que l'enfermement affirmé dans une école de pensée. Le terme *pragmatisme* se réfère en effet à différents courants qui s'opposent sur la plupart des points (Tiercelin, 2002). Je trouve dans Putnam, Hacking ou Searle des points de convergence qui permettent de définir cette méthode pour la psychologie ; c'est-à-dire une méthode non-relativiste mais bien réaliste, permettant d'appréhender les objets de la psychologie avec un dynamisme/renégociation qui me paraît absolument nécessaire.

La méthode que je propose pour appréhender les propriétés mentales rentre en contradiction avec la démarche usuelle en science telle que pratiquée par les lignes éditoriales des revues académiques en psychologie. Ces revues imposent un mimétisme, une inertie, absolument délétère pour la science. Donc la démarche proposée dans cette thèse me paraît difficilement applicable de façon concrète pour un chercheur désireux de publier, car renégocier les concepts et mesures d'une façon pratique rentre en contradiction avec le carcan méthodologique usuel. Raison pour laquelle je considère que les travaux critiques des mesures en psychologie ayant été publiés dans des revues académiques, même s'ils me paraissent mal poser les problèmes, sont salutaires et ouvrent des fenêtres nécessaires car mettant un coin dans la pensée dominante, coin sans lequel je n'aurais pas pu construire ma thèse.

Le pragmatisme comme critique de la pratique de la science

Je suis de ceux qui considèrent que la science ne devrait pas avoir pour objectif d'apporter uniquement des réponses techniques, mais d'être inscrite avant tout dans une démarche critique, à l'instar de ce que Dewey (1925/2012) dit de la philosophie « *la philosophie apparaît alors comme une théorie généralisée de la critique* ». L'objectif de la science ne devrait pas être orienté que vers l'opérationnalisation de ses développements car comme le dit Putnam (1984) « *aucune personne cultivée ne dira que la seule valeur de la science, c'est ses applications pratiques* » (p. 199) ; mais bien inscrire la science dans le processus de connaissance et de leur critique. Je renvoie pour exemple à la critique de « *l'argument*

scientifique » de Gould (1997) sur la mesure de l'homme fondamentalement inscrit au XIX^{ème} siècle dans une idéologie raciste et eugéniste porté par la quasi-unanimité des scientifiques de l'époque. Nous sommes non seulement légitimes d'avoir une démarche critique identique de la production scientifique contemporaine, mais cette démarche critique est absolument impérative pour construire cette démarche scientifique. Et cette critique doit exister en psychologie.

Comme je l'ai montré (et en phase avec bien d'autres auteurs), la production psychométrique peut être regardée avec une certaine critique. Si les psychologues n'entendent le plus souvent pas cette critique, c'est que la psychométrie a une fonction sociale et politique qui entraîne les acteurs de ce champ scientifique à épouser la dynamique sociale, à apporter un crédit scientifique à des mesures dont la fonction normative sert de levier à des orientations politiques. J'ai à de nombreuses reprises critiqué les auteurs porteurs de la critique de la mesure (Michell, Vautier, Maraun entre autres). Si je suis critique de leurs travaux c'est que je les considère inscrits dans une épistémologie empiriste-réaliste critiquable. Mais a contrario je considère que leurs travaux sont salutaires. Leurs développements nourrissent la pensée critique nécessaire de la pratique usuelle de la mesure en psychologie, et, au-delà, la démarche critique nécessaire de la pensée scientifique. Il faut en effet affirmer un processus de démocratisation de la Science « *La crise des sciences est aussi une crise de l'expertise dont les solutions sont en partie connues : il faut démocratiser l'expertise, diffuser à tous un maximum d'éléments de compréhension des enjeux et des solutions mises en place et redonner au citoyen une place dans la décision en s'appuyant à la fois sur les éléments de la démocratie formelle [...] et sur des expériences de démocratie participative* » (Armatte, 2010, p. 13).

Comme le dit A. Gramin (in Desrosières, 2008b), on mesure tout et n'importe quoi en psychologie, sans pudeur, sans recul, sans question. La littérature académique offre un panel de mesures absolument incroyable sans que soit souvent questionnée cette pratique. La personne se résume à une mosaïque de scores qui la représente (Lamiell, 2000). Si la psychologie n'interroge pas sa pratique, c'est que ses acteurs ne sont guère différents des autres acteurs sociaux, dominés par la pression sociale et la dynamique de leur espace social (ne serait-ce que par la pression bibliométrique qui oblige de répliquer ce qui se fait pour publier et interdit le plus souvent toute démarche critique). Autant je considère potentiel de mesurer des propriétés mentales comme je l'ai développé ; autant je rejoins la critique dans l'usage sociale de la mesure et la critique de sa pseudo-scientificité. On fait de la *psycho*

technique comme le dit Vautier, inscrit dans des enjeux sociaux de normalisation et hiérarchisation qui ne sont pas indépendantes d'un cadre politique.

Cette critique de la psychométrie doit se comprendre comme critique de sa fonction sociale et politique. La psychométrie use des méthodes quantitatives comme « une évidence » (je reprends les termes de O. Martin), sans regard critique, pour imiter les sciences de la nature par des modèles basés sur un discours logico-mathématique. La quantification permet à la psychologie d'intégrer le cercle fermé des « sciences ». La démarche psychométrique devient dès lors une production de mesures sans possibilité de critique pour l'espace public. Ces mesures encadrent les individus, où la singularité n'est appréhendée qu'au regard d'une norme. Je rejoins donc Hacking dans sa critique de la fonction politique de la psychologie. S'appuyant sur Foucault, Hacking considère que le bio-pouvoir, qui d'après Foucault cherche à obtenir « *l'assujettissement des corps et le contrôle de la population* » (Hacking, 2006, p. 337), peut s'étendre à « l'esprit » : « *En devenant avec Ebbinghaus une science statistique, [la psychologie] s'est désintéressée des événements ou des êtres individuels pour ne plus s'intéresser qu'aux moyennes et aux déviations. Elle fit alors partie du bio-pôle généralisé [...] qui, en fait, donne une image fidèle de l'essence des "contrôle régulateurs" de Foucault* ».

Il faut, comme le font certains auteurs, affirmer que la pratique de la mesure doit rompre avec une expertise autoritaire des scientifiques et inscrire cette pratique dans une démarche de réappropriation par l'espace public de ces mesures. Rompre avec l'affirmation de l'exactitude de ces mesures et affirmer une pratique scientifique critique, inscrite dans un rapport social, demandant en retour la critique des personnes mesurées.

Mon travail s'inscrit donc dans cette démarche critique qui me paraît être l'essence de la démarche scientifique. Critique de la mesure telle que pratiquée ; critique de la critique de la mesure pour ses soubassements épistémologiques.

Pragmatisme et critique sociale

Comme je l'ai dit en introduction, le pragmatisme renvoie à une démarche intellectuelle critique. Hacking (2008) considère que la démarche philosophique posant que les objets de la science sont en partie construits renvoie en effet à une position critique au sens social ; lui-même se définissant comme « *positivist and anarchist* » (Hacking, 2007). C'est parce que le pragmatisme offre un cadre philosophique critique, que j'ai été puiser dans ce courant ce qui

me paraissait impératif de considérer en psychologie afin de répondre aux contradictions de toute la littérature académique.

Ma position est critique de la méthode dominante en psychologie, et, au-delà, critique de l'usage social des mesures psychologiques, et de façon plus fondamentale de la fonction politique de ces mesures et du rôle des scientifiques dans cette fonction d'autorité morale. Développer une critique rentre nécessairement dans le champ politique. Le courant qui a été dominant dans la critique sociale est le marxisme. Le pragmatisme est un courant philosophique clairement en rupture avec une forme de matérialisme scientiste et positiviste tel que l'a développé entre autre tout un courant marxiste (voir la critique du marxisme dans Castoriadis, 1999). Cependant je m'autorise à considérer que ce que je développe ne me semble pas en contradiction avec une certaine lecture de Marx... mais donc en contradiction avec la lecture de Marx que font Putnam (1984) ou Pereira (2010) par exemple, mais aussi de la lecture de Marx de beaucoup de marxistes (la liste est longue de Engels à Bensaid). On peut trouver des critiques du scientisme marxiste chez des auteurs se réclamant par ailleurs d'un *open marxism*, comme Holloway par exemple, c'est-à-dire puisant dans la méthode développée par Marx des ressources critiques sans adhérer au marxisme orthodoxe, et renvoyant à un cadre de la connaissance proche de ce j'ai développé. Holloway (2008, p. 196) écrit en effet « *il est impossible de dire "j'ai une connaissance de la réalité" simplement parce que les catégories au moyen desquelles chacun l'appréhende sont des catégories historiquement spécifiques faisant partie de cette réalité* ».

Je considère que le processus d'objectivation de la réalité est une praxis, amenant une objectivité subjectivée par une pratique au sens social ; et ceci est parfaitement en phase avec Marx où le réel est appréhendé en tant « *qu'activité humaine sensible, en tant que praxis, de façon subjective* » (Première thèse sur Feuerbach, Marx & Engels, 1846/1972). Dans sa critique du matérialisme du début du XIXème siècle, Marx considère qu'il faut appréhender les objets réels par l'activité objective de l'homme, c'est l'activité humaine qui rend objectif la réalité car l'homme est avant tout « *un produit social* » pour Marx : « *l'homme est dans le sens le plus littéral un animal politique, non seulement un animal sociable, mais un animal qui ne peut s'isoler que dans la société* » (p. 150). Ainsi, pour Marx : « *le principal défaut de tout matérialisme jusqu'ici [...] est que l'objet extérieur, la réalité, le sensible ne sont saisis que sous la forme d'Objet ou d'intuition, mais non en tant qu'activité humaine sensible, en tant que praxis, de façon subjective [...] Feuerbach veut des objets sensibles, réellement distincts des objets de la pensée : mais il ne saisit pas l'activité humaine elle-même en tant qu'activité objective* » (Première thèse sur Feuerbach, Marx & Engels, 1846/1972).

Nous retrouvons toujours dans la seconde thèse sur Feuerbach la considération de la praxis comme activité de connaissance et comme rapport à la « vérité ». Ici la « vérité » est ramenée à la « réalité pratique », c'est donc bien une réalité objectivée par la pratique : « *La question de savoir si la pensée humaine peut aboutir à une vérité objective n'est pas une question théorique, mais une question pratique. C'est dans la pratique qu'il faut que l'homme prouve la vérité, c'est-à-dire la réalité [...] de sa pensée* » (Seconde thèse sur Feuerbach, Marx & Engels, 1846/1972). De même le langage est un « produit social », il y a donc bien un filtre (Putnam)/arrière-plan (Searle)/matrice (Hacking) chez Marx dans sa compréhension du processus de conscientisation : « *Le langage est aussi vieux que la conscience, le langage est la conscience réelle, pratique, existant aussi pour d'autres hommes, existant donc alors seulement pour moi-même aussi [...] La conscience est donc d'emblée un produit social et le demeure aussi longtemps qu'il existe des hommes* » (Marx & Engels, 1846/1972, p.63).

Mais fondamentalement, comme je l'ai déjà dit, je peux faire mien ce cadre posé par Marx sur la connaissance de la réalité, en gardant à l'esprit la l'objectivité pratique (donc subjectivée) que Marx a posé, Marx (1859/1972) « *Le concret est concret parce qu'il est la synthèse de multiple déterminations, donc unité de la diversité. C'est pourquoi il apparaît dans la pensée comme procès de synthèse, comme résultat, non comme point de départ, bien qu'il soit le véritable point de départ et par la suite également le point de départ de la vue immédiate et de la représentation [...] la méthode qui consiste à s'élever de l'abstrait au concret n'est pour la pensée que la manière de s'approprier le concret, de le reproduire sous la forme de concret pensé. Mais ce n'est nullement là le procès de la genèse du concret lui-même [...] Pour la conscience le mouvement des catégories apparaît comme l'acte de production réel – qui reçoit une simple impulsion du dehors et on le regrette – dont le résultat est le monde ; et ceci (mais c'est encore là une tautologie) est exact dans la mesure où la totalité concrète en tant que totalité pensée, en tant que représentation mentale du concret, est en fait un produit de la pensée, de la conception ; il n'est par contre nullement le produit du concept qui s'engendrerait lui-même, qui penserait en dehors et au-dessus de la vue immédiate et de la représentation, mais un produit de l'élaboration de concepts à partir de la vue immédiate et de la représentation [...] Après comme avant, le sujet réel subsiste dans son indépendance en dehors de l'esprit [...] Par conséquent, dans l'emploi de la méthode théorique aussi, il faut que le sujet, la société, reste constamment présent à l'esprit comme donnée première* » (p. 165-166).

L'objectif n'est pas d'encapsuler mon développement dans une pensée marxiste, car je ne cherche pas à me référer à une école de pensée close sur elle-même. Mon objectif, par ce

détour par Marx, est bien de positionner la pensée pragmatique que je défends dans une forme de pensée critique sociale plus globale. Considérer l'activité de connaissance, et donc de recherche, comme une pratique inscrite dans une matrice sociale renvoie à la critique de cette matrice pour développer des activités de recherche en tant qu'activité sociale, et donc politique. Et je crois que mon travail s'inscrit dans cette dynamique critique, critique de la forme et fonction de la psychométrie actuelle ; et constructive, dans ses propositions d'alternatives démocratisantes des mesures en psychologie, en considérant qu'une mesure est une praxis, inscrite dans un dialogue, nécessitant la transparence, et devant refuser de se considérer comme « valide » ou « normative » au sens où la psychométrie l'entend le plus souvent.

Bibliographie

- Aboulafia, M., & Shook, J. R. (2008). *Contemporary Pragmatism: Volume 5, Number 1. June 2008*.
Rodopi.
- Ahuvia, A., Bagozzi, R. P., & Batra, R. (2013). Psychometric vs. C-oar-se measures of brand love: A
reply to rossiter. *Marketing Letters*. <http://doi.org/10.1007/s11002-013-9251-4>
- Allen, P. M., & Clough, S. (2015). Philosophical commitments, empirical evidence, and theoretical
psychology. *Theory & Psychology*, 25(1), 3–24. <http://doi.org/10.1177/0959354314563324>
- Allouche, J.-P. (2014, January 29). Connaître n'est pas mesurer. *Le Monde*. Retrieved from
<http://www.scoop.int/nouveaux-paradigmes/p/4015029386/2014/01/29/connaitre-n-est-pas-mesurer>
- Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral
and Brain Sciences*, 33(04), 245–266. <http://doi.org/10.1017/S0140525X10000853>
- Andrieu, B. (2007). La chair vécue du cerveau : un objet épistémologique du cerveau psychologique.
*Contributions de L'épistémologie, de La Philosophie Des Sciences et de La Psychologie Aux
Sciences Cognitives*, 52(3), 315–325. <http://doi.org/10.1016/j.psfr.2007.04.001>
- Armatte, M. (2005). La notion de modèle dans les sciences sociales: anciennes et nouvelles
significations. *Mathématiques et sciences humaines*, (172). <http://doi.org/10.4000/msh.2962>
- Armatte, M. (2010). *La science économique comme ingénierie : Quantification et Modélisation*.
Presses de l'Ecole des mines.
- Avenier, M. J. (2011). Pourquoi jeter le bébé avec l'eau du bain ? Méthodologie sans épistémologie
n'est que ruine de la réflexion ! Retrieved from <http://halshs.archives-ouvertes.fr/halshs-00642673>
- Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective
measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*,
12(2), 229–237. <http://doi.org/10.1037/1082-989X.12.2.229>

- Bagozzi, R. P. (2011a). Alternative Perspectives in Philosophy of Mind and Their Relationship to Structural Equation Models in Psychology. *Psychological Inquiry*, 22(2), 88–99.
<http://doi.org/10.1080/1047840X.2010.550183>
- Bagozzi, R. P. (2011b). Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations. *MIS Quarterly*, 35(2), 261–292.
- Bailleux, C., Dauvier, B., & Perret, P. (2015). L'augmentation des capacités en mémoire de travail est-elle une condition nécessaire, bien que non suffisante, au développement de l'intelligence ? In *Différences et variabilités en psychologie*. P.U.R.
- Bainter, S. A., & Bollen, K. A. (2014). Interpretational Confounding or Confounded Interpretations of Causal Indicators? *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 125–140. <http://doi.org/10.1080/15366367.2014.968503>
- Bainter, S. A., & Bollen, K. A. (2015). Moving Forward in the Debate on Causal Indicators: Rejoinder to Comments. *Measurement: Interdisciplinary Research and Perspectives*, 13(1), 63–74.
<http://doi.org/10.1080/15366367.2015.1016349>
- Bamber, D., & van Santen, J. P. . (2000). How to Assess a Model's Testability and Identifiability. *Journal of Mathematical Psychology*, 44(1), 20–40. <http://doi.org/10.1006/jmps.1999.1275>
- Barandiaran, X., & Moreno, A. (2008). On the nature of neural information: A critique of the received view 50 years later. *Neural Networks: Algorithms and Applications 4th International Symposium on Neural Networks 50 Years of Artificial Intelligence: A Neuronal Approach Campus Multidisciplinary in Perception and Intelligence*, 71(4–6), 681–692.
<http://doi.org/10.1016/j.neucom.2007.09.014>
- Barberousse, Kistler, & Ludwig. (2000). *La philosophie des sciences au XXè siècle* (Flamarion). Paris.
- Barrett, L. F. (2009a). The Future of Psychology: Connecting Mind to Brain. *Perspectives on Psychological Science*, 4(4), 326–339. <http://doi.org/10.1111/j.1745-6924.2009.01134.x>

- Barrett, L. F. (2009b). Understanding the Mind by Measuring the Brain: Lessons From Measuring Behavior (Commentary on Vul et al., 2009). *Perspectives on Psychological Science*, 4(3), 314–318. <http://doi.org/10.1111/j.1745-6924.2009.01131.x>
- Barrett, L. F. (2011). Bridging Token Identity Theory and Supervenience Theory Through Psychological Construction. *Psychological Inquiry*, 22(2), 115–127. <http://doi.org/10.1080/1047840X.2011.555216>
- Barrett, P. (2005). What If There Were No Psychometrics?: Constructs, Complexity, and Measurement. *Journal of Personality Assessment*, 85(2), 134–140. http://doi.org/10.1207/s15327752jpa8502_05
- Barrett, P. (2008). The Consequence of Sustaining a Pathology: Scientific Stagnation— a Commentary on the Target Article “Is Psychometrics a Pathological Science?” by Joel Michell. *Measurement*, 6(1/2), 78–83. <http://doi.org/10.1080/15366360802035521>
- Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson’s bonds model of intelligence. *Psychological Review*, 116(3), 567–579. <http://doi.org/10.1037/a0016262>
- Bartholomew, & Knott. (1999). *Latent variable models and factor analysis*. Londres: Arnold.
- Bedau, M. A. (1997). Weak emergence. *Philosophical Perspectives*, 11, 375–379.
- Bedau, M. A. (2008). Is Weak Emergence Just in the Mind? *Minds & Machines*, 18(4), 443–459. <http://doi.org/10.1007/s11023-008-9122-6>
- Belis, M. (1995). Causalité, propension, probabilité -. *Intellectica*, 21(2), 199–231.
- Bem, S., & Jong, H. L. de. (2006). *Theoretical Issues in Psychology: An Introduction*. SAGE.
- Bennett, Baird, Miller, & Wolford. (2011). Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction. *Journal of Serendipitous and Unexpected Results*, 1(1), 1–5.
- Bentler, P. M. (1986). Structural modeling and psychometrika: An historical perspective on growth and achievements. *Psychometrika*, 51(1), 35–51. <http://doi.org/10.1007/BF02293997>

- Berkman, E. T., & Lieberman, M. D. (2011). What's Outside the Black Box?: The Status of Behavioral Outcomes in Neuroscience Research. *Psychological Inquiry*, 22(2), 100–107.
<http://doi.org/10.1080/1047840X.2011.550182>
- Berthoz, A. (2013). *La vicariance : le cerveau créateur de mondes*. O. Jacob. Retrieved from
[http://babordplus.univ-bordeaux.fr/notice.php?q=titre_col%3A\(%22Odile%20Jacob%20sciences%22\)&spec_expand=&start=0](http://babordplus.univ-bordeaux.fr/notice.php?q=titre_col%3A(%22Odile%20Jacob%20sciences%22)&spec_expand=&start=0)
- Bertrand, D., Ahmadi, A. E., & Heuchenne, C. (2008). D'une échelle ordinale de Guttman à une échelle de rapports de Rasch. *Mathématiques et sciences humaines*, (184), 25–46.
<http://doi.org/10.4000/msh.10959>
- Bertrand, D., Ahmadi, A. E., & Heuchenne, C. (2009). Mathématiques et sciences humaines - Mathematics and social sciences. *Mathématiques et sciences humaines. Mathematics and social sciences*, (184), 25–46. <http://doi.org/10.4000/msh.10959>
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley.
- Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, 12(2), 219–228.
<http://doi.org/10.1037/1082-989X.12.2.219>
- Bollen, K. A. (2011). Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models. *MIS Quarterly*, 35(2), 359–372.
- Bollen, K. A., & Arminger. (1991). Observational residuals in factor analysis and structural equation models. *Sociological Methodology*, 21, 235–262.
- Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychological Methods*, 16(3), 265–284.
<http://doi.org/10.1037/a0024448>

- Bollen, K. A., & Davis, W. R. (2009a). Causal Indicator Models: Identification, Estimation, and Testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 498–522.
<http://doi.org/10.1080/10705510903008253>
- Bollen, K. A., & Davis, W. R. (2009b). Two Rules of Identification for Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 523–536.
<http://doi.org/10.1080/10705510903008261>
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305–314. <http://doi.org/10.1037/0033-2909.110.2.305>
- Bontly, T. D. (2002). The Supervenience Argument Generalizes. *Philosophical Studies*, 109(1), 75–96.
- Boomsma, A. (2013). Reporting Monte Carlo Studies in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(3), 518–540.
<http://doi.org/10.1080/10705511.2013.797839>
- Borsboom, D. (2008a). A Tour Guide to the Latent Realm. *Measurement*, 6(1/2), 134–146.
<http://doi.org/10.1080/15366360802121941>
- Borsboom, D. (2008b). Latent Variable Theory. *Measurement*, 6(1/2), 25–53.
<http://doi.org/10.1080/15366360802035497>
- Borsboom, D. (2012). Whose Consensus Is It Anyway? Scientific Versus Legalistic Conceptions of Validity. *Measurement: Interdisciplinary Research & Perspective*, 10(1-2), 38–41.
<http://doi.org/10.1080/15366367.2012.681971>
- Borsboom, D. (2015). What is causal about individual differences? : A comment on Weinberger. *Theory & Psychology*, 25(3), 362–368. <http://doi.org/10.1177/0959354315587784>
- Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, 9(1), 91–121.
<http://doi.org/10.1146/annurev-clinpsy-050212-185608>

- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franic, S. (2009). The end of construct validity. In *The concept of validity* (IAP, pp. 135–172).
- Borsboom, D., Kievit, R. A., Cervone, D., & Hood, S. B. (2009). The Two Disciplines of Scientific Psychology, or: The Disunity of Psychology as a Working Hypothesis. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudhary (Eds.), *Dynamic Process Methodology in the Social and Developmental Sciences* (pp. 67–97). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-0-387-95922-1_4
- Borsboom, D., & Markus, K. A. (2013). Truth and Evidence in Validity Theory: Truth and Evidence in Validity Theory. *Journal of Educational Measurement*, *50*(1), 110–114.
<http://doi.org/10.1111/jedm.12006>
- Borsboom, D., & Mellenbergh, G. J. (2004). Why Psychometrics is Not Pathological A Comment on Michell. *Theory & Psychology*, *14*(1), 105–120. <http://doi.org/10.1177/0959354304040200>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. <http://doi.org/10.1037/0033-295X.110.2.203>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. <http://doi.org/10.1037/0033-295X.111.4.1061>
- Borsboom, D., & Scholten, A. Z. (2008). The Rasch Model and Conjoint Measurement Theory from the Perspective of Psychometrics. *Theory & Psychology*, *18*(1), 111–117.
<http://doi.org/10.1177/0959354307086925>
- Bourdieu, P. (1997). *Les usages sociaux de la science : Pour une sociologie clinique du champ* (Quae). INRA.
- Bourdieu, P. (2001). *Science de la science et réflexivité* (Raisons d’agir). Paris: Raisons d’agir.
- Boyd, danah, & Crawford, K. (2011). Six Provocations for Big Data. *SSRN Electronic Journal*.
<http://doi.org/10.2139/ssrn.1926431>

- Brennan, R. L. (2013). Commentary on “Validating the Interpretations and Uses of Test Scores”:
 Brennan Commentary. *Journal of Educational Measurement*, 50(1), 74–83.
<http://doi.org/10.1111/jedm.12001>
- Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F.
 (2013). A Network Approach to Psychopathology: New Insights into Clinical Longitudinal
 Data. *PLoS ONE*, 8(4), 1–13. <http://doi.org/10.1371/journal.pone.0060188>
- Brown, S. D. (2012). Common ground for behavioural and neuroimaging research. *Australian Journal
 of Psychology*, 64(1), 4–10. <http://doi.org/10.1111/j.1742-9536.2011.00046.x>
- Burnston, D., Sheredos, B., & Bechtel, W. (2011). HIT on the Psychometric Approach. *Psychological
 Inquiry*, 22(2), 108–114. <http://doi.org/10.1080/1047840X.2011.547387>
- Burt, R. S. (1976). Interpretational Confounding of Unobserved Variables in Structural Equation
 Models. *Sociological Methods & Research*, 5(1), 3–52.
<http://doi.org/10.1177/004912417600500101>
- Cadogan, J. W., & Lee, N. (2013). Improper use of endogenous formative variables. *Journal of
 Business Research*, 66(2), 233–241. <http://doi.org/10.1016/j.jbusres.2012.08.006>
- Cadogan, J. W., Souchon, A. L., & Procter, D. B. (2008). The quality of market-oriented behaviors:
 Formative index construction. *Formative Indicators Formative Indicators*, 61(12), 1263–1277.
<http://doi.org/10.1016/j.jbusres.2008.01.014>
- Callon, M., & Latour, B. (1982). *La science telle qu'elle se fait*. La Découverte.
- Cariani, P. (1997). Emergence of new signal-primitives in neural systems. *Intellectica*, 25(2), 95–143.
- Carnino, G. (2015). *L'Invention de la science : La nouvelle religion de l'âge industriel*. Paris: Seuil.
- Castoriadis, C. (1999). *L'institution imaginaire de la société*. Seuil.
- Cavalla, C., & Grossmann, F. (2004). Semantic characteristics and use of some specific cross-
 disciplinary lexemes in French scientific papers. In *Colloque international Comparing forms
 academic organisation, cultural identities and scientific discourse across nations and
 disciplines*. Paris.

- Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of Formative Measurement in Information Systems Research. *MIS Quarterly*, 33(4), 689–707.
- Cervone, D., & Caldwell, T. L. (2008). From Measurement Theory to Psychological Theory, in Reverse. *Measurement*, 6(1/2), 84–88. <http://doi.org/10.1080/15366360802035539>
- Charlton, R. A., Landau, S., Schiavone, F., Barrick, T. R., Clark, C. A., Markus, H. S., & Morris, R. G. (2008). A structural equation modeling investigation of age-related variance in executive function and DTI measured white matter damage. *Neurobiology of Aging*, 29(10), 1547–1555. <http://doi.org/10.1016/j.neurobiolaging.2007.03.017>
- Chauveau, S. (2014). Science, industrie, innovation et société au XIX^e siècle. *Le Mouvement Social*, 248(3), 3. <http://doi.org/10.3917/lms.248.0003>
- Chauviré, C. (2010). La philosophie comme description de l'ordinaire chez Peirce et chez Wittgenstein. *Archives de Philosophie*, 73(1), 81–91.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research*, 36(4), 462–494. <http://doi.org/10.1177/0049124108314720>
- Chiao, & Cheon. (2010). The weirdest brains in the world. Retrieved from http://culturalneuro.psych.northwestern.edu/ChiaoCheon_BBS_inpress.pdf
- Chin, W. W. (1998, March). Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, pp. 1–1.
- Chin, W. W. (2010). How to write up and report PLS analyses. In *Handbook of Partial Least Squares* (V. Esposito Vinzi, W. W. Chin, J. Henseler, H. Wang, pp. 655–690). Londres: Springer.
- Chin, W. W., Thatcher, J. B., & Wright, R. T. (2012). Assessing Common Method Bias: Problems with the Ulmc Technique. *MIS Quarterly*, 36(3), 1003–A11.
- Cho, S.-J., Cohen, A. S., Kim, S.-H., & Bottge, B. (2010). Latent Transition Analysis With a Mixture Item Response Theory Measurement Model. *Applied Psychological Measurement*, 34(7), 483–504. <http://doi.org/10.1177/0146621610362978>

- Churchill, J. R., & Gilbert, A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research (JMR)*, 16(1), 64–73.
- Churchill, J. R., Peter, J. P., & Gilbert, A. (1984). Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research (JMR)*, 21(4), 360–375.
- Cliff. (1993). MEASUREMENT THEORY AND THE REVOLUTION THAT NEVER HAPPENED., 3(3), 186–190.
- Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Formative Indicators Formative Indicators*, 61(12), 1250–1262. <http://doi.org/10.1016/j.jbusres.2008.01.013>
- Cometti, J.-P. (2001). Le pluralisme pragmatiste et la question du relativisme. *Archives de Philosophie*, 64(1), 21–39.
- Cometti, J.-P. (2010). *Qu'est-ce que le pragmatisme ?* Paris: Folio.
- Comte, A. (1830). *Cours de philosophie positive, 1re et 2e leçons*. Paris. Retrieved from http://classiques.uqac.ca/classiques/Comte_auguste/cours_philo_positive/cours_philo_positive.html
- Conner, C. D. (2011). *Histoire populaire des sciences*. Montreuil: Editions L'échappée.
- Corcuff, P. (2002). De Rosa Luxemburg à la social-démocratie libertaire. *Contre Temps*, (6), 101–108.
- Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... Borsboom, D. (2012). Dimensions of Normal Personality as Networks in Search of Equilibrium: You Can't Like Parties if You Don't Like People. *European Journal of Personality*, 26(4), 414–431. <http://doi.org/10.1002/per.1866>
- Cromby, J., & Willis, M. E. H. (2014). Nudging into subjectification: Governmentality and psychometrics. *Critical Social Policy*, 34(2), 241–259. <http://doi.org/10.1177/0261018313500868>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <http://doi.org/10.1037/h0040957>

- Danziger, K. (1985). The Methodological Imperative in Psychology. *Philosophy of the Social Sciences*, 15(1), 1–13. <http://doi.org/10.1177/004839318501500101>
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. ., & Starr, J. M. (2000). The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, 28(1), 49–55. [http://doi.org/10.1016/S0160-2896\(99\)00031-8](http://doi.org/10.1016/S0160-2896(99)00031-8)
- Debreu, G. (1959). *Topological Methods in Cardinal Utility Theory* (Cowles Foundation Discussion Paper No. 76). Cowles Foundation for Research in Economics, Yale University. Retrieved from <https://ideas.repec.org/p/cwl/cwldpp/76.html>
- De Ribaupierre, A. (2015). Pourquoi faut-il étudier la variabilité intra-individuelle ? In *Différences et variabilités en psychologie* (pp. 159–178). P.U.R.
- Desrosières, A. (2008a). *Gouverner par les nombres*. Paris: Presses de l’Ecole des mines.
- Desrosières, A. (2008b). *Pour une sociologie historique de la quantification*. Paris: Presses de l’Ecole des mines.
- Desrosières, A. (2010). *La politique des grands nombres*. Paris: LA DECOUVERTE.
- Desrosières, A. (2014). *Prouver et gouverner*. Paris: La Découverte.
- Dewey, J. (1925). *Expérience et nature*. Gallimard.
- Diallo, T. M. O., & Morin, A. J. S. (2015). Power of Latent Growth Curve Models to Detect Piecewise Linear Trajectories. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 449–460. <http://doi.org/10.1080/10705511.2014.935678>
- Diamantopoulos, A. (2005). The C-OAR-SE procedure for scale development in marketing: A comment. *International Journal of Research in Marketing*, 22(1), 1–9. <http://doi.org/10.1016/j.ijresmar.2003.08.002>
- Diamantopoulos, A. (2010). Reflective and formative metrics of relationship value: Response to Baxter’s commentary essay. *Journal of Business Research*, 63(1), 91–93. <http://doi.org/10.1016/j.jbusres.2009.03.001>

- Diamantopoulos, A. (2011). Incorporating Formative Measures into Covariance-Based Structural Equation Models. *MIS Quarterly*, 35(2), 335–358.
- Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203–1218.
<http://doi.org/10.1016/j.jbusres.2008.01.009>
- Diamantopoulos, A., & Sigauw, J. A. (2006). Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *British Journal of Management*, 17(4), 263–282. <http://doi.org/10.1111/j.1467-8551.2006.00500.x>
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research (JMR)*, 38(2), 269–277.
- Dickes, P., Tournaois, J., Flieller, A., & Kop, J.-L. (1994). *Psychométrie*. PUF.
- Dietvorst, R. C., Verbeke, W. J. M. ., Bagozzi, R. P., Yoon, C., Smits, M., & van der Lugt, A. (2009). A Sales Force–Specific Theory-of-Mind Scale: Tests of Its Validity by Classical Methods and Functional Magnetic Resonance Imaging. *Journal of Marketing Research (JMR)*, 46(5), 653–668. <http://doi.org/10.1509/jmkr.46.5.653>
- Drouet, I. (2012). Propensions popperiennes et puissances aristotéliennes: De l’interprétation des probabilités à la métaphysique. *Philosophie*, 114(2), 50.
<http://doi.org/10.3917/philo.114.0050>
- Durieux, P., Bruxelles, J., Savignoni, A., & Coste, J. (2001). Prévalence et prise en charge de la douleur à l’hôpital. *La Presse Médicale*, 30(12), 572–576.
- Edwards, J. R. (2011). The Fallacy of Formative Measurement. *Organizational Research Methods*, 14(2), 370–388. <http://doi.org/10.1177/1094428110378369>
- Edwards J.R., & Bagozzi R.P. (2000). On the Nature and Direction of Relationships Between Constructs and Measures. *Psychological Methods*, 5(2), 155–174.
- El-Hani, C., & Pihlström, S. (2002). Emergence Theories and Pragmatic Realism. *Essays in Philosophy*, 3(2). Retrieved from <http://commons.pacificu.edu/eip/vol3/iss2/3>

- Enders, G., Enders, J., & Liber, I. (2015). *Le charme discret de l'intestin : Tout sur un organe mal aimé*. Arles: Actes Sud Editions.
- Engelhard, G. (2008). Historical Perspectives on Invariant Measurement: Guttman, Rasch, and Mokken. *Measurement*, 6(3), 155–189. <http://doi.org/10.1080/15366360802197792>
- Engels, F. (1888). *Ludwig feuerbach et la fin de la philosophie classique allemande* (Editions Sociales). Editions Sociales.
- Falissard, B. (2008a). *Cerveau et psychanalyse, tentative de réconciliation* (L'harmattan). Paris.
- Falissard, B. (2008b). *Mesurer la subjectivité en santé : Perspective méthodologique et statistique* (2e édition). Issy-les-Moulineaux: Elsevier Masson.
- Falissard, B. (2012). *Le « sujet » de la santé entre science, psychanalyse et philosophie*. Presented at the Rencontre philisophique d'Uriage, Uriage. Retrieved from <https://www.youtube.com/watch?v=0bOEe2BsDLw>
- Falissard, B. (2014, July). *Le sujet de la rencontre*. Presented at the IXe journée de phénoménologie de Nice, Nice. Retrieved from https://www.youtube.com/watch?v=juFYEm_F9fo
- Falissard, B., Révah, A., Yang, S., & Fagot-Largeault, A. (2013). The place of words and numbers in psychiatric research. *Philosophy, Ethics, and Humanities in Medicine*, 8(1), 1–9. <http://doi.org/10.1186/1747-5341-8-18>
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56–83. <http://doi.org/10.1080/10705519909540119>
- Feyerabend, P. (1975). *CONTRE LA METHODE. Esquisse d'une théorie anarchiste de la connaissance*. Paris: Seuil.
- Feyerabend, P. (1979). *ecrits philosophiques t.1 ; réalisme rationalisme et methode scientifique*. DIANOIA.
- Fine, A. (2001). The scientific image twenty years later. *Philosophical Studies*, 106(1-2), 107 – 122.

- Finkelstein, L. (2003). Widely, strongly and weakly defined measurement. *Measurement*, 34(1), 39–48. [http://doi.org/10.1016/S0263-2241\(03\)00018-6](http://doi.org/10.1016/S0263-2241(03)00018-6)
- Finkelstein, L. (2005). Problems of measurement in soft systems. *Measurement*, 38(4), 267–274. <http://doi.org/10.1016/j.measurement.2005.09.002>
- Finn, A., & Kayande, U. (1997). Reliability Assessment and Optimization of Marketing Measurement. *Journal of Marketing Research (JMR)*, 34(2), 262–275.
- Finn, A., & Kayande, U. (2005). How fine is C-OAR-SE? A generalizability theory perspective on Rossiter's procedure. *International Journal of Research in Marketing*, 22(1), 11–21. <http://doi.org/10.1016/j.ijresmar.2004.03.001>
- Finn, A., & Wang, L. (2014). Formative vs. reflective measures: Facets of variation. *Journal of Business Research*, 67(1), 2821–2826. <http://doi.org/10.1016/j.jbusres.2012.08.001>
- Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass: A Bradford Book / MIT Press.
- Fornell, C., & Bookstein, F. L. (1982). Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. *Journal of Marketing Research (JMR)*, 19(4), 440–452.
- Fornell, C., & Cha, J. (1994). Partial least squares. In *Advanced methods of marketing research* (Blackwell, pp. 52–78). Cambridge, England: Bagozzi R. P.
- Foucart, J. (2014). Pragmatisme et transaction. La perspective de John Dewey. *Pensée plurielle*, 33-34(2), 73–84. <http://doi.org/10.3917/pp.033-034.0073>
- Francis, J. E. (2009). Is C-OAR-SE best for internet retailing service quality? *Managing Service Quality*, 19(6), 670–686. <http://doi.org/10.1108/09604520911005062>
- Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research*, 61(12), 1229–1237. <http://doi.org/10.1016/j.jbusres.2008.01.011>
- Gardner, H. (2008). *Les intelligences multiples*. Paris: Retz.

- Geert, P. van. (1994). *Dynamic systems of development : change between complexity and chaos / Paul van Geert*. Hemel Hempstead: Harvester Wheatsheaf.
- Gefen, D., Rigdon, E. E., & Straub, D. (2011). An Update and Extension to SEM Guidelines for Administrative and Social Science Research. *MIS Quarterly*, 35(2), iii–A7.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Comparing Pls to Regression and Lisrel: A Response to Marcoulides, Chin, and Saunders. *MIS Quarterly*, 36(3), 703–A10.
- Gori, R. (2014). *Les citoyens contre la bureaucratie des évaluations*. Presented at the RIGP 2014. Retrieved from http://www.dailymotion.com/video/x205dh0_rigp-2014-les-citoyens-contre-la-bureaucratie-des-evaluations_school
- Gould, S. J. (1997). *La Mal-mesure de l'homme* (Nouv éd rev. et augm). Paris: O. Jacob.
- Granger, G.-G. (2002). *Formes opérations objets*. Paris: Librairie Philosophique Vrin.
- Grattan, K. T. V., Sun, T., & Khan, S. H. (2013). Ludwik Finkelstein and measurement – A challenge for the future. *Measurement*, 46(8), 2990–2992. <http://doi.org/10.1016/j.measurement.2013.04.053>
- Green, S. B., & Yang, Y. (2009). Commentary on Coefficient Alpha: A Cautionary Tale. *Psychometrika*, 74(1), 121–135. <http://doi.org/10.1007/s11336-008-9098-4>
- Gureckis, T. M., James, T. W., & Nosofsky, R. M. (2011). Re-evaluating dissociations between implicit and explicit category learning: An event-related fMRI study. *Journal of Cognitive Neuroscience*, 23(7), 1697–1709. <http://doi.org/10.1162/jocn.2010.21538>
- Guttman, L. (1944). A Basis for Scaling Qualitative Data. *American Sociological Review*, 9(2), 139–150.
- Guyon, H., & Tensaout, M. (2015). Formative Measurement Models: A Response to Bainter & Bollen (2014) and Howell (2014). *Measurement: Interdisciplinary Research and Perspectives*, 13(1), 53–58. <http://doi.org/10.1080/15366367.2015.1016340>
- Hacking, I. (1992). Statistical language, statistical truth and statistical reason: The self-authentication of a style of reasoning. In *Social Dimensions of Science* (pp. 130–157). University of Notre Dame Press.

- Hacking, I. (2002). *L'Emergence de la Probabilité*. Paris: Seuil.
- Hacking, I. (2006). *L'AME REECRITE. Etude sur la personnalité multiple et les sciences de la mémoire*.
Les empêcheurs de tourner en rond.
- Hacking, I. (2007). On Not Being a Pragmatist : Eight Reasons and a Cause. In C. J. Misak (Ed.), *New Pragmatists* (p. 32). Oxford University Press.
- Hacking, I. (2008). *Entre science et réalité*. Paris: LA DECOUVERTE.
- Haertel, E. (2013). Getting the Help We Need: Getting the Help We Need. *Journal of Educational Measurement*, 50(1), 84–90. <http://doi.org/10.1111/jedm.12002>
- Haig, B. D. (2005a). An Abductive Theory of Scientific Method. *Psychological Methods*, 10(4), 371–388.
- Haig, B. D. (2005b). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, 40(3), 303–329.
- Haig, B. D. (2008a). Aspects of Latent Variable Theory. *Measurement*, 6(1/2), 88–93.
<http://doi.org/10.1080/15366360802035547>
- Haig, B. D. (2008b). Précis of “An abductive theory of scientific method.” *Journal of Clinical Psychology*, 64(9), 1019–1022.
- Haig, B. D. (2008c). Scientific method, abduction, and clinical reasoning. *Journal of Clinical Psychology*, 64(9), 1013–1018.
- Haig, B. D. (2011). Philosophical Naturalism and Scientific Method. *Psychological Inquiry*, 22(2), 128–136. <http://doi.org/10.1080/1047840X.2011.552055>
- Haig, B. D., & Borsboom, D. (2008). On the Conceptual Foundations of Psychological Measurement. *Measurement*, 6(1/2), 1–6. <http://doi.org/10.1080/15366360802035471>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2013). *A Primer on Partial Least Squares Structural Equation Modeling*. Los Angeles: SAGE Publications Inc.
- Hanafi, M. (2007). PLS Path modelling: computation of latent variables with the estimation mode B. *Computational Statistics*, 22(2), 275–292. <http://doi.org/10.1007/s00180-007-0042-3>

- Harari-Kermadec, H., & Bruno, I. (2012). Le benchmarking et les enjeux politiques de la quantification. Presented at the Les pensées critiques contemporaines , EHESS. Retrieved from <http://pcc.hypotheses.org/63>
- Hardin, A., Chang, J. C.-J., & Fuller, M. (2008). Formative vs. Reflective Measurement: Comment on Marakas, Johnson, and Clay (2007). *Journal of the Association for Information Systems*, 9(9). Retrieved from <http://aisel.aisnet.org/jais/vol9/iss9/3>
- Hardin, A., Chang, J. C.-J., Fuller, M. A., & Torkzadeh, G. (2011). Formative Measurement and Academic Research: In Search of Measurement Theory. *Educational and Psychological Measurement*, 71(2), 281–305. <http://doi.org/10.1177/0013164410370208>
- Hardin, A., & Marcoulides, G. A. (2011). A Commentary on the Use of Formative Measurement. *Educational and Psychological Measurement*, 71(5), 753–764. <http://doi.org/10.1177/0013164411414270>
- Harley, T. A. (2004). Does Cognitive Neuropsychology Have a Future? *Cognitive Neuropsychology*, 21(1), 3–16. <http://doi.org/10.1080/02643290342000131>
- Heeger, D. J., & Ress, D. (2002). WHAT DOES fMRI TELL US ABOUT NEURONAL ACTIVITY? *Nature Reviews Neuroscience*, 3(2), 142–151. <http://doi.org/10.1038/nrn730>
- Heene, M. (2013). Additive conjoint measurement and the resistance toward falsifiability in psychology. *Quantitative Psychology and Measurement*, 246. <http://doi.org/10.3389/fpsyg.2013.00246>
- Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics*, 25(1), 107–120. <http://doi.org/10.1007/s00180-009-0164-x>
- Herrnstein, R. J. Murray, Charles, & Murray, C. (1994). *Bell Curve: Intelligence and Class Structure in American Life*. Free Press.
- Holloway, J. (2008). *Changer le monde sans prendre le pouvoir : Le sens de la révolution aujourd'hui*. Paris; Montréal: Editions Syllepse.

- Hood, S. B. (2008). Comments on Borsboom's Typology of Measurement Theoretic Variables and Michell's Assessment of Psychometrics as "Pathological Science." *Measurement*, 6(1/2), 93–97. <http://doi.org/10.1080/15366360802035554>
- Hood, S. B. (2012). In Defense of an Instrument-Based Approach to Validity. *Measurement: Interdisciplinary Research & Perspective*, 10(1-2), 63–65. <http://doi.org/10.1080/15366367.2012.681976>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling An Overview and a Meta-Analysis. *Sociological Methods & Research*, 26(3), 329–367. <http://doi.org/10.1177/0049124198026003003>
- Howell, R. D. (2008). Observed Variables are Indeed More Mysterious than Commonly Supposed. *Measurement*, 6(1/2), 97–101. <http://doi.org/10.1080/15366360802121826>
- Howell, R. D. (2014). What is the Latent Variable in Causal Indicator Models? *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 141–145. <http://doi.org/10.1080/15366367.2014.980105>
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods*, 12(2), 238–245. <http://doi.org/10.1037/1082-989X.12.2.238>
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2007b). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205–218. <http://doi.org/10.1037/1082-989X.12.2.205>
- Howell R.D., Breivik E., & Wilcox J.B. (2007). REPLY: Is Formative Measurement Really Measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods*, 12(2), 238–245.
- Huetz, C. (2006, January 1). *Etude du code neuronal sous-tendant la perception de signaux de communication acoustique*. Paris 11. Retrieved from <http://www.theses.fr/2006PA112293>
- Humphreys, P. (1985). Why Propensities Cannot be Probabilities. *Philosophical Review*, 94(4), 557–570.
- Humphreys, P. (1997a). Emergence, not supervenience. *Philosophy of Science*, 64, 337–345.

- Humphreys, P. (1997b). How properties emerge. *Philosophy of Science*, 64(1), 1.
- Humphreys, P. (2008). Computational and Conceptual Emergence. *Philosophy of Science*, 75(5), 584–594.
- Hyatt, S., & Collins, L. (2000). Inference in Latent Transition Models: results from a simulation study.
- Jarvis, C. B., Mackenzie, S. B., Podsakoff, P. M., Mick, D. G., & Bearden, W. O. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, 30(2), 199–218.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International.
- Jöreskog, K. G., Sörbom, D., & Wallentin, F. Y. (2006). *Latent Variable Scores and Observational Residuals*.
- Juhel, J. (2005). La psychométrie : la recherche et l'évaluation des compétences qui caractérisent l'intelligence (L'intelligence, pp. 23–39). Hermès Science.
- Juhel, J. (2013a). La recherche d'invariants différentiels dans les variations développementales : de la population à l'individu et réciproquement ! In *Développements et variabilités* (P.U.R).
- Juhel, J. (2013b). *Niveaux d'observation et approches méthodologiques dans l'étude de la variabilité intra-individuelle*. Presented at the Journée scientifique "la dépression à travers le temps : entre variabilité et stabilité," Nantes.
- Juhel, J. (2014). *Où est l'individuel dans l'étude des différences individuelles de personnalité?* Presented at the Journée d'étude "La personnalité : double éclairage différentiel et psychosocial," Rouen.
- Juhel, J. (2015). Modèles structuraux et inférence causale. In *Différences et variabilités en psychologie* (pp. 309–352). P.U.R. Retrieved from http://www.academia.edu/14157193/Juhel_J._2015_.Mod%C3%A8les_structuraux_et_inf%C3%A9rence_causale._In_J._Juhel_and_G._Rouxel_Eds._Diff%C3%A9rences_et_variabilit%C3%A9s_en_psychologie_pp._309-352._Rennes_P.U.R

- Juhel, J., & Gilles, P.-Y. (2011). *VALIDITÉ DES MODÈLES ET DES OUTILS DE L'EXAMEN PSYCHOLOGIQUE*. Presented at the Conférence de consensus en psychologie.
- Juhel, J., & Rouxel, G. (2015). *Différences et variabilités en psychologie* (P.U.R.). Retrieved from <http://livre.fnac.com/a7880988/Jacques-Juhel-Differences-et-variabilites-en-psychologie>
- Junker, B. W., & Sijtsma, K. (2001). Nonparametric Item Response Theory in Action: An Overview of the Special Issue. *Applied Psychological Measurement, 25*(3), 211–220.
<http://doi.org/10.1177/01466210122032028>
- Kagan, J. (2005). A Time for Specificity. *Journal of Personality Assessment, 85*(2), 125–127.
http://doi.org/10.1207/s15327752jpa8502_03
- Kane, M. (2008). The Benefits and Limitations of Formality. *Measurement, 6*(1/2), 101–108.
<http://doi.org/10.1080/15366360802035562>
- Kane, M. (2013a). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement, 50*(1), 1–73. <http://doi.org/10.1111/jedm.12000>
- Kane, M. (2013b). Validation as a Pragmatic, Scientific Activity. *Journal of Educational Measurement, 50*(1), 115–122. <http://doi.org/10.1111/jedm.12007>
- Karabatsos, G., & Sheu, C.-F. (2004). Order-Constrained Bayes Inference for Dichotomous Models of Unidimensional Nonparametric IRT. *Applied Psychological Measurement, 28*(2), 110–125.
<http://doi.org/10.1177/0146621603260678>
- Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness: replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular Psychiatry, 17*(4), 377–388. <http://doi.org/10.1038/mp.2011.182>
- Ke, Z., & Wang, L. (Peggy). (2015). Detecting Individual Differences in Change: Methods and Comparisons. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(3), 382–400.
<http://doi.org/10.1080/10705511.2014.936096>

- Khan, S., & Finkelstein, L. (2013). Mathematical modelling in the analysis and design of hard and soft measurement systems. *Measurement*, 46(8), 2936–2941.
<http://doi.org/10.1016/j.measurement.2013.04.045>
- Kievit, R. A., Romeijn, J.-W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011a). Mind the Gap: A Psychometric Approach to the Reduction Problem. *Psychological Inquiry*, 22(2), 67–87. <http://doi.org/10.1080/1047840X.2011.550181>
- Kievit, R. A., Romeijn, J.-W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011b). Modeling Mind and Matter: Reductionism and Psychological Measurement in Cognitive Neuroscience. *Psychological Inquiry*, 22(2), 139–157.
<http://doi.org/10.1080/1047840X.2011.567962>
- Kievit, R. A., van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience*, 3(2), 89–97. <http://doi.org/10.1080/17588928.2011.628383>
- Kim, G., Shin, B., & Grover, V. (2010). Investigating Two Contradictory Views of Formative Measurement in Information Systems Research. *MIS Quarterly*, 34(2), 345–A5.
- Kim, J. (1997). Explanation, Prediction, and Reduction in Emergentism. *Intellectica*, 25(2), 45–57.
- Kim, J. (1999). Making sens of emergence. *Philosophical Studies*, 95, 3–36.
- Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, 151(3), 547–559.
<http://doi.org/10.1007/s11229-006-9025-0>
- Kim, J. (2009). *La survenance et l'esprit*. Paris: Ithaque.
- Kim, S.-Y. (2012). Sample Size Requirements in Single- and Multiphase Growth Mixture Models: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3), 457–476. <http://doi.org/10.1080/10705511.2012.687672>
- Kistler, M. (2007). La réduction, l'émergence, l'unité de la science et les niveaux de réalité. *Matière Première*, (2), 67–97.

- Kline, R. (2010). *Principles and Practice of Structural Equation Modeling* (third). New York: GUILFORD Press.
- Koo, N., & Leite, W. L. (2014). The Impact of Ignoring Time Series Processes in Linear Growth Mixture Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 210–224.
<http://doi.org/10.1080/10705511.2014.882669>
- Körner, S. (1962). On Empirical Continuity: *Monist*, 47(1), 1–19.
<http://doi.org/10.5840/monist19624711>
- Körner, S. (1964). Science and Moral Responsibility. *Mind*, 161–172.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (2006a). *Foundations of Measurement Volume I: Additive and Polynomial Representations*. Dover Publications.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (2006b). *Foundations of Measurement Volume II: Geometrical, Threshold, and Probabilistic Representations*. Dover Publications.
- Kuhn, T.-S. (1970). *La structure des révolutions scientifiques*. Paris: Flammarion.
- Kuncel, N. R., Ones, D. S., & Hezlett, S. A. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological Bulletin*, 127(1), 162–181.
- Kyngdon, A. (2008a). Conjoint Measurement, Error and the Rasch Model A Reply to Michell, and Borsboom and Zand Scholten. *Theory & Psychology*, 18(1), 125–131.
<http://doi.org/10.1177/0959354307086927>
- Kyngdon, A. (2008b). The Rasch Model from the Perspective of the Representational Theory of Measurement. *Theory & Psychology*, 18(1), 89–109.
<http://doi.org/10.1177/0959354307086924>
- Kyngdon, A. (2008c). Treating the Pathology of Psychometrics: An Example from the Comprehension of Continuous Prose Text. *Measurement*, 6(1/2), 108–113.
<http://doi.org/10.1080/15366360802035570>

- Kyngdon, A. (2011a). Plausible measurement analogies to some psychometric models of test performance. *British Journal of Mathematical & Statistical Psychology*, 64(3), 478–497.
<http://doi.org/10.1348/2044-8317.002004>
- Kyngdon, A. (2011b). Psychological Measurement Needs Units, Ratios, and Real Quantities: A Commentary on Humphry. *Measurement*, 9(1), 55–58.
<http://doi.org/10.1080/15366367.2011.558791>
- Kyngdon, A. (2013). Descriptive theories of behaviour may allow for the scientific measurement of psychological attributes. *Theory & Psychology*, 23(2), 227–250.
<http://doi.org/10.1177/0959354312468221>
- Lacot, E., Afzali, M. H., & Vautier, S. (2015). Test Validation Without Measurement. *European Journal of Psychological Assessment*, 1–11. <http://doi.org/10.1027/1015-5759/a000253>
- Laird, A., Lancaster, J., & Fox, P. (2005). BrainMap. *Neuroinformatics*, 3(1), 65–77.
- Lakatos, I. (1970). *Criticism and the Growth of Knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Cambridge University Press.
- Lamiell, J. T. (2000). A periodic table of personality elements? The “Big Five” and trait “psychology” in critical perspective. *Journal of Theoretical and Philosophical Psychology*, 20(1), 1–24.
<http://doi.org/10.1037/h0091211>
- Lamiell, J. T. (2006). La psychologie contemporaine des « traits » dans le cadre de la recherche néogaltonienne : comment elle est censée fonctionner et pourquoi en réalité elle ne fonctionne pas. *Psychologie Française*, 51(3), 337–355.
<http://doi.org/10.1016/j.psfr.2005.08.001>
- Lamiell, J. T. (2013). Statisticism in personality psychologists’ use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology*, 31(1), 65–71.
<http://doi.org/10.1016/j.newideapsych.2011.02.009>

- Lanza, S. T., & Collins, L. M. (2008). A New SAS Procedure for Latent Transition Analysis: Transitions in Dating and Sexual Risk Behavior. *Developmental Psychology*, 44(2), 446–456.
<http://doi.org/10.1037/0012-1649.44.2.446>
- Lecerf, T., & Kieng, S. (2015). WISC-IV : bref état de l'art. In *Différences et variabilités en psychologie* (pp. 19–36). P.U.R.
- Lee, N., Cadogan, J. W., & Chamberlain, L. (2013). The MIMIC model and formative variables: problems and solutions. *AMS Review*, 3(1), 3–17. <http://doi.org/10.1007/s13162-013-0033-1>
- Lektorskii, V. A. (2010). Realism, Antirealism, Constructivism, and Constructive Realism in Contemporary Epistemology and Science. *Journal of Russian & East European Psychology*, 48(6), 5–44. <http://doi.org/10.2753/RPO1061-0405480601>
- Le Moigne, J.-L. (2001). Pourquoi je suis un constructiviste non repentant. *Revue du MAUSS*, 17(1), 197–223. <http://doi.org/10.3917/rdm.017.0197>
- Le Moigne, J.-L. (2010). *Les épistémologies constructivistes* (3e éd). Paris: [diffusion] Cairn.info.
- Leplège, A. (2001). *Méthodologie des mesures de qualité de vie en médecine: aspects épistémologiques et éthiques*.
- Leplège, A. (2003a). Editorial. Epistemology of Measurement in the Social Sciences: Historical and Contemporary Perspectives. *Social Science Information*, 42(4), 451–462.
<http://doi.org/10.1177/0539018403424001>
- Leplège, A. (2003b). Note Technique: Invariance et Modèles de Mesure dans les Sciences Sociales, Introduction à l'œuvre de Georg Rasch. *Social Science Information*, 42(4), 535–555.
<http://doi.org/10.1177/0539018403424005>
- Leplège, A., & Coste, J. (2002). *Mesure de la santé perceptuelle et de la qualité de vie : méthodes et applications*. Paris: Editions Estem.
- Lissitz, R. W. (2009). Introduction. In *The concept of validity* (IAP, pp. 1–18).
- Lissitz, R. W. (2010). *The concept of validity*. Information Age Publishing.

- Lissitz, R. W., & Caliço, T. (2012). Validity is an Action Verb: Commentary on: “Clarifying the Consensus Definition of Validity.” *Measurement: Interdisciplinary Research & Perspective*, 10(1-2), 75–79. <http://doi.org/10.1080/15366367.2012.677346>
- Lovasz, & Slaney. (2013). What makes a hypothetical construct “hypothetical”? Tracing the origins and uses of the ‘hypothetical construct’ concept in psychological science. *New Ideas in Psychology*, 31(1), 22–31.
- Luce, R. D. (1977). Thurstone’s discriminial processes fifty years later. *Psychometrika*, 42(4), 461–489.
- Luce, R. D. (1997). Several Unresolved Conceptual Problems of Mathematical Psychology. *Journal of Mathematical Psychology*, 41(1), 79–87. <http://doi.org/10.1006/jmps.1997.1150>
- Luce, R. D., & Suppes, P. (2002). Representational Measurement Theory. In H. Pashler (Ed.), *Stevens’ Handbook of Experimental Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc. Retrieved from <http://doi.wiley.com/10.1002/0471214426.pas0401>
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1), 1–27.
- Ludwig, P. (2004). Le Problème du corps et de l’esprit. Approches contemporaines. *Coté Philo*, 4, 34–40.
- Lytard, J.-F. (1979). *LA CONDITION POSTMODERNE. : Rapport sur le savoir* (Les éditions De Minuit). Paris: Editions de Minuit.
- MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114(3), 533–541. <http://doi.org/10.1037/0033-2909.114.3.533>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <http://doi.org/10.1037/1082-989X.1.2.130>

- MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions. *Journal of Applied Psychology, 90*(4), 710–730.
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in Mis and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly, 35*(2), 293–A5.
- Madelrieux, S. (2012). Méthode ou métaphysique ? L'empirisme pragmatique de John Dewey. *Critique, 787*(12), 1043–1058.
- Marakas, G., Johnson, R., & Clay, P. F. (2008). Formative vs. Reflective Measurement: A Reply to Hardin, Chang, and Fuller. *Journal of the Association for Information Systems, 9*(9). Retrieved from <http://aisel.aisnet.org/jais/vol9/iss9/2>
- Maraun, M. D. (2010). *Myths and confusions: Psychometrics and the latent variable model*. Retrieved from <http://www.sfu.ca/~maraun/Mikes%20page-%20Myths%20and%20Confusions.html>
- Maraun, M. D. (2012). Validity and Measurement. *Measurement: Interdisciplinary Research & Perspective, 10*(1-2), 80–83. <http://doi.org/10.1080/15366367.2012.682523>
- Maraun, M. D., & Gabriel. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas in Psychology, 31*(1), 33–42.
- Maraun, M. D., & Halpin, P. F. (2008). Manifest and Latent Variates. *Measurement, 6*(1/2), 113–117. <http://doi.org/10.1080/15366360802035596>
- Maraun, M. D., & Peters, J. (2005). What Does It Mean That an Issue Is Conceptual in Nature? *Journal of Personality Assessment, 85*(2), 128–133. http://doi.org/10.1207/s15327752jpa8502_04
- Maraun, M. D., Slaney, & Gabriel. (2005). The Augustinian methodological family of psychology. *New Ideas in Psychology, 27*, 148–162.
- Marcoulides, G. A., & Chin, W. (2013). You Write, but Others Read: Common Methodological Misunderstandings in PLS and Related Methods. In H. Abdi, W. W. Chin, V. Esposito Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New Perspectives in Partial Least Squares and Related*

- Methods* (Vol. 56, pp. 31–64). Springer New York. Retrieved from
http://dx.doi.org/10.1007/978-1-4614-8283-3_2
- Marcoulides, G. A., Chin, W. W., & Saunders, C. (2009). A Critical Look at Partial Least Squares Modeling. *MIS Quarterly*, *33*(1), 171–175.
- Mari, L. (2013). A quest for the definition of measurement. *Measurement*, *46*(8), 2889–2895.
<http://doi.org/10.1016/j.measurement.2013.04.039>
- Mari, L., Carbone, P., & Petri, D. (2012). Measurement Fundamentals: A Pragmatic View. *IEEE Transactions on Instrumentation and Measurement*, *61*(8), 2107–2115.
<http://doi.org/10.1109/TIM.2012.2193693>
- Markus, K. A. (2008a). Constructs, Concepts and the Worlds of Possibility: Connecting the Measurement, Manipulation, and Meaning of Variables. *Measurement*, *6*(1/2), 54–77.
<http://doi.org/10.1080/15366360802035513>
- Markus, K. A. (2008b). Putting Concepts and Constructs into Practice: A Reply to Cervone and Caldwell, Haig, Kane, Mislevy, and Rupp. *Measurement*, *6*(1/2), 147–154.
<http://doi.org/10.1080/15366360802122006>
- Markus, K. A. (2012). Constructs and Attributes in Test Validity: Reflections on Newton’s Account. *Measurement: Interdisciplinary Research & Perspective*, *10*(1-2), 84–87.
<http://doi.org/10.1080/15366367.2012.677348>
- Markus, K. A. (2014). Unfinished Business in Clarifying Causal Measurement: Commentary on Bainter and Bollen. *Measurement: Interdisciplinary Research and Perspectives*, *12*(4), 146–150.
<http://doi.org/10.1080/15366367.2014.980106>
- Markus, K. A., & Borsboom, D. (2012). The cat came back: Evaluating arguments against psychological measurement. *Theory & Psychology*, *22*(4), 452–466.
<http://doi.org/10.1177/0959354310381155>
- Markus, K. A., & Borsboom, D. (2013a). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY, US: Routledge/Taylor & Francis Group.

- Markus, K. A., & Borsboom, D. (2013b). Reflective measurement models, behavior domains, and common causes. *New Ideas in Psychology, 31*(1), 54–64.
<http://doi.org/10.1016/j.newideapsych.2011.02.008>
- Martin, J. (2003). Positivism, Quantification and the Phenomena of Psychology. *Theory & Psychology, 13*(1), 33–38. <http://doi.org/10.1177/0959354303013001760>
- Martin, O. (1997). *La mesure de l'esprit. Origines et développements de la psychometrie, 1900-1950*. L'harmattan.
- Marx, K. (1859). *Contribution à la critique de l'économie politique*. Paris: Edition Sociale.
- Marx, K., & Engels, F. (1846). *L'idéologie allemande*. Les Editions sociales.
- Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22*(2), 89–113. [http://doi.org/10.1016/S0160-2896\(96\)90011-2](http://doi.org/10.1016/S0160-2896(96)90011-2)
- McGrane, J. A. (2015). Stevens' forgotten crossroads: the divergent measurement traditions in the physical and psychological sciences from the mid-twentieth century. *Quantitative Psychology and Measurement, 431*. <http://doi.org/10.3389/fpsyg.2015.00431>
- McGrath, R. E. (2005a). Conceptual Complexity and Construct Validity. *Journal of Personality Assessment, 85*(2), 112–124. http://doi.org/10.1207/s15327752jpa8502_02
- McGrath, R. E. (2005b). Rethinking Psychosocial Constructs: Reply to Comments by Barrett, Kagan, and Maraun and Peters. *Journal of Personality Assessment, 85*(2), 141–145.
http://doi.org/10.1207/s15327752jpa8502_06
- McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Networks, 13*(8–9), 861–870.
[http://doi.org/10.1016/S0893-6080\(00\)00059-9](http://doi.org/10.1016/S0893-6080(00)00059-9)
- McLaughlin, B. P. (2008). Emergence and supervenience. In M. A. Bedau & P. Humphreys (Eds.), *Emergence: Contemporary readings in philosophy and science*. (pp. 81–97). Cambridge, MA US: MIT Press.

Meehl, P. E. (1990). WHY SUMMARIES OF RESEARCH ON PSYCHOLOGICAL THEORIES ARE OFTEN UNINTERPRETABLE. *Psychological Reports*, 66(1), 195–244.

<http://doi.org/10.2466/pr0.1990.66.1.195>

Merleau-Ponty, M. (1960). *Signes*. Folio.

Michell. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology*, 31(1), 13–21.

Michell, J. (1997a). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–383.

Michell, J. (1997b). Reply to Kline, Laming, Lovie, Luce, and Morgan. *British Journal of Psychology*, 88(3), 401.

Michell, J. (1999). *Measurement in psychology* (Cambridge university press). Cambridge, England.

Michell, J. (2000). Normal Science, Pathological Science and Psychometrics. *Theory & Psychology*, 10(5), 639–667. <http://doi.org/10.1177/0959354300105004>

Michell, J. (2003a). Pragmatism, Positivism and the Quantitative Imperative. *Theory & Psychology*, 13(1), 45–52. <http://doi.org/10.1177/0959354303013001761>

Michell, J. (2003b). The Quantitative Imperative Positivism, Naive Realism and the Place of Qualitative Methods in Psychology. *Theory & Psychology*, 13(1), 5–31.
<http://doi.org/10.1177/0959354303013001758>

Michell, J. (2004). Item Response Models, Pathological Science and the Shape of Error Reply to Borsboom and Mellenbergh. *Theory & Psychology*, 14(1), 121–129.
<http://doi.org/10.1177/0959354304040201>

Michell, J. (2005). The Meaning of the Quantitative Imperative A Response to Niaz. *Theory & Psychology*, 15(2), 257–263. <http://doi.org/10.1177/0959354305051369>

Michell, J. (2008a). Conjoint Measurement and the Rasch Paradox A Response to Kyngdon. *Theory & Psychology*, 18(1), 119–124. <http://doi.org/10.1177/0959354307086926>

- Michell, J. (2008b). Is Psychometrics Pathological Science? *Measurement*, 6(1/2), 7–24.
<http://doi.org/10.1080/15366360802035489>
- Michell, J. (2008c). Rejoinder. *Measurement*, 6(1/2), 125–133.
<http://doi.org/10.1080/15366360802121917>
- Michell, J. (2009). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical & Statistical Psychology*, 62(1), 41–55. <http://doi.org/10.1348/00071107X243582>
- Michell, J. (2012). "The constantly recurring argument": Inferring quantity from order. *Theory & Psychology*, 22(3), 255–271. <http://doi.org/10.1177/0959354311434656>
- Miller, G. A. (2010). Mistreating Psychology in the Decades of the Brain. *Perspectives on Psychological Science (Sage Publications Inc.)*, 5(6), 716–743. <http://doi.org/10.1177/1745691610388774>
- Misak, C. J. (2007). *New Pragmatists*. Oxford University Press.
- Mislevy, R. J. (2008). How Cognitive Science Challenges the Educational Measurement Tradition. *Measurement*, 6(1/2), 124–124. <http://doi.org/10.1080/15366360802131635>
- Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In *The concept of validity* (IAP, pp. 83–108).
- Molenaar, P. C. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement*, 2(4), 201–218.
- Moss, P. A. (2013). Validity in Action: Lessons From Studies of Data Use: Validity in Action. *Journal of Educational Measurement*, 50(1), 91–98. <http://doi.org/10.1111/jedm.12003>
- Nagel, E. (1961). *The structure of science*.
- Needham, P. (2009). Reduction and emergence: a critique of Kim. *Philosophical Studies*, 146(1), 93–116. <http://doi.org/10.1007/s11098-008-9246-9>
- Newby, V. A., Conner, G. R., Grant, C. P., & Bunderson, C. V. (2009). The Rasch model and additive conjoint measurement. *Journal Of Applied Measurement*, 10(4), 348–354.
- Newton, P. E. (2012). Clarifying the Consensus Definition of Validity. *Measurement: Interdisciplinary Research & Perspective*, 10(1-2), 1–29. <http://doi.org/10.1080/15366367.2012.669666>

- Newton, P. E. (2013). Two Kinds of Argument?: Two Kinds of Argument? *Journal of Educational Measurement, 50*(1), 105–109. <http://doi.org/10.1111/jedm.12004>
- Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research, 12*(5), 419. <http://doi.org/10.1191/0962280203sm341ra>
- Nicodème, F. (2015). Marx, ou comment être à la fois matérialiste et révolutionnaire. Retrieved from <http://www.franceculture.fr/player/reecouter?play=5058109>
- Noaparast, K. B., & Khosravi, Z. (2006). Mind and Mental Health Based on a Realistic Constructivism. *Constructivism in the Human Sciences, 11*(1-2), 20–31.
- Osborne, J. W. (2010). Challenges for quantitative psychology and measurement in the 21st century. *Frontiers in Psychology, 1*(1). <http://doi.org/10.3389/fpsyg.2010.00001>
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo Experiments: Design and Implementation. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(2), 287–312. http://doi.org/10.1207/S15328007SEM0802_7
- Peirce, C. S. (1879). *La Logique de la science*. Editions la Bibliothèque Digitale.
- Pelagatti, M. M., Fattore, M., & Vittadini, G. (2012). Inconsistencies of the PLS-PM approach to structural equation models with formative-reflective schemes. *Electronic Journal of Applied Statistical Analysis, 5*(3), 333–338. <http://doi.org/10.1285/i20705948v5n3p333>
- Pereira, I. (2009, June 3). *Un nouvel esprit contestataire – La grammaire pragmatiste du syndicalisme d'action directe libertaire* (phdthesis). Ecole des Hautes Etudes en Sciences Sociales (EHESS). Retrieved from <https://tel.archives-ouvertes.fr/tel-00392699/document>
- Pereira, I. (2010). *Peut-on être radical et pragmatique ?* Paris: Textuel.
- Pereira, I. (2012). POUR UN USAGE ANARCHISTE DU PRAGMATISME Irène Pereira. Retrieved August 31, 2015, from <http://1libertaire.free.fr/AnarchismePragmatisme01.html>
- Peter, J. P. (1979). Reliability: A Review of Psychometric Basics and Recent Marketing Practices. *Journal of Marketing Research (JMR), 16*(1), 6–17.

- Petitmengin, C. (2009). The Validity of First-Person Descriptions as Authenticity and Coherence. *Journal of Consciousness Studies*, 16(10-12), 10–12.
- Petter, S., Straub, D., & Rai, A. (2007). Specifying Formative Constructs in Information Systems Research. *MIS Quarterly*, 31(4), 623–656.
- Peugh, J., & Fan, X. (2013). Modeling Unobserved Heterogeneity Using Latent Profile Analysis: A Monte Carlo Simulation. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 616–639. <http://doi.org/10.1080/10705511.2013.824780>
- Peugh, J., & Fan, X. (2015). Enumeration Index Performance in Generalized Growth Mixture Models: A Monte Carlo Test of Muthén's (2003) Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 115–131. <http://doi.org/10.1080/10705511.2014.919823>
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59–63. <http://doi.org/10.1016/j.tics.2005.12.004>
- Poldrack, R. A. (2008). The role of fMRI in Cognitive Neuroscience: where do we stand? *Current Opinion in Neurobiology*, 18(2), 223–227. <http://doi.org/10.1016/j.conb.2008.07.006>
- Poldrack, R. A. (2010). Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? *Perspectives on Psychological Science*, 5(6), 753–761. <http://doi.org/10.1177/1745691610388777>
- Popper, K. (1979). Logique Des Sciences Sociales. In *De Vienne à Francfort, la querelle allemande des sciences sociales* (pp. 75–90). Editions Complexes. Retrieved from <https://fr.scribd.com/doc/50782286/Popper-Logique-Des-Sciences-Sociales>
- Popper, K. (1984). *POST-SCRIPTUM A LA LOGIQUE DE LA DECOUVERTE SCIENTIFIQUE. Tome 2, L'Univers irrésolu, Plaidoyer pour l'indéterminisme*. Paris: Editions Hermann.
- Prosser, S. (2012). Emergent Causation. *Philosophical Studies*, 159(1), 21–39. <http://doi.org/10.1007/s11098-010-9685-y>
- Putnam, H. (1984). *Raison, vérité et histoire*. Editions de Minuit.
- Putnam, H. (1998). *Philosophie de la logique*. Combas: Editions de l'Eclat.

- Putnam, H. (2004). *Fait/Valeur : la fin d'un dogme et autres essais*. Paris: Editions de l'Eclat.
- Putnam, H. (2011). *Le Réalisme à visage humain*. Paris: Gallimard.
- Quéré, L. (2001). Naturaliser le sens : une erreur de catégorie ? *Revue du MAUSS*, 17(1), 275.
<http://doi.org/10.3917/rdm.017.0275>
- Qureshi, I., & Compeau, D. (2009). Assessing Between-Group Differences in Information Systems Research: A Comparison of Covariance- and Component-Based Sem. *MIS Quarterly*, 33(1), 197–214.
- Raykov, T., & Penev, S. (2013). Exploring Structural Equation Model Misspecification via Latent Individual Residuals. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent Variable and Latent Structure Models* (1 edition, pp. 121–134). New York u.a.: Psychology Press.
- Rentz, J. O. (1987). Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures. *Journal of Marketing Research (JMR)*, 24(1), 19–28.
- Rhemtulla, M., Bork, R. van, & Borsboom, D. (2015). Calling Models With Causal Indicators “Measurement Models” Implies More Than They Can Deliver. *Measurement: Interdisciplinary Research and Perspectives*, 13(1), 59–62. <http://doi.org/10.1080/15366367.2015.1016343>
- Rigdon, E. E. (2012). Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods. *Long Range Planning*, 45(5–6), 341–358. <http://doi.org/10.1016/j.lrp.2012.09.010>
- Rigdon, E. E. (2014). Comment on “Improper use of endogenous formative variables.” *Journal of Business Research*, 67(1), 2800–2802. <http://doi.org/10.1016/j.jbusres.2012.08.005>
- Rigdon, E. E., Preacher, K. J., Lee, N., Howell, R. D., Franke, G. R., & Borsboom, D. (2011). Avoiding measurement dogma: A response to Rossiter. *European Journal of Marketing*, 45(11-12), 1589–1600. <http://doi.org/10.1108/03090561111167306>
- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, 49(1), 37–47. <http://doi.org/10.1007/BF02294204>

- Ringle, C. M., Götz, O., Wetzels, M., & Wilson, B. (2009). On the Use of Formative Measurement Specifications in Structural Equation Modeling: A Monte Carlo Simulation Study to Compare Covariance-Based and Partial Least Squares Model Estimation Methodologies [MPRA Paper]. Retrieved October 8, 2013, from <http://mpra.ub.uni-muenchen.de/15390/>
- Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A Critical Look at the Use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, 36(1), iiv–8.
- Riou, J., Guyon, H., & Falissard, B. (2015). An introduction to the partial least squares approach to structural equation modelling: a method for exploratory psychiatric research. *International Journal of Methods in Psychiatric Research*. <http://doi.org/10.1002/mpr.1497>
- Rossiter, J. R. (2002). The C-OAR-SE Procedure For Scale Development in Marketing. *International Journal of Research in Marketing*, 19, 305–355.
- Rossiter, J. R. (2011). Marketing measurement revolution: The C-OAR-SE method and why it must replace psychometrics. *European Journal of Marketing*, 45(11-12), 1561–1588. <http://doi.org/10.1108/03090561111167298>
- Rossiter, J. R. (2012). A new C-OAR-SE-based content-valid and predictively valid measure that distinguishes brand love from brand liking. *Marketing Letters*, 23(3), 905–916. <http://doi.org/10.1007/s11002-012-9173-6>
- Rossiter, J. R., & Braithwaite, B. (2013). C-OAR-SE-based single-item measures for the two-stage Technology Acceptance Model. *Australasian Marketing Journal (AMJ)*, 21(1), 30–35.
- Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research*, 20(4), 235–249. <http://doi.org/10.1002/mpr.352>
- Rouxel, G. (2015). L’approche des traits a-t-elle un avenir ? In *Différences et variabilités en psychologie* (pp. 125–149). P.U.R.
- Rozencwajg, P. (2015). Pour une approche intégrative de l’intelligence : le cas des cubes de Kohs. In *Différences et variabilités en psychologie* (pp. 63–81). P.U.R.

- Rupp, A. A. (2008). Lost in Translation? Meaning and Decision Making in Actual and Possible Worlds. *Measurement*, 6(1/2), 117–123. <http://doi.org/10.1080/15366360802035612>
- Sadin, E. (2015). *La vie algorithmique : Critique de la raison numérique*. Paris: L'ECHAPPEE EDITIONS.
- Salzberger, T. (2013). Attempting Measurement of Psychological Attributes. *Frontiers in Psychology*, 4. <http://doi.org/10.3389/fpsyg.2013.00075>
- Scheemaekere, X. D. (2009). *Les fondements philosophiques du concept de probabilité* (Working Papers CEB No. 09-025.RS). ULB -- Université Libre de Bruxelles. Retrieved from <https://ideas.repec.org/p/sol/wpaper/09-025.html>
- Schinckus, C. (2008). Chapitre 5. Constructivisme et relativisme ou l'expression de la postmodernité dans les sciences sociales. In *Épistémologie de la sociologie* (pp. 97–120). De Boeck Supérieur. Retrieved from http://www.cairn.info/article.php?ID_ARTICLE=DBU_JACQU_2008_02_0097
- Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, 31(1), 43–53. <http://doi.org/10.1016/j.newideapsych.2011.02.007>
- Scholten, A. Z. (2011a). The Guttman-Rasch paradox in item response theory. In *Admissible statistics from a latent variable perspective* (pp. 47–72).
- Scholten, A. Z. (2011b). The Rasch model as additive conjoint measurement. In *Admissible statistics from a latent variable perspective* (p. 171).
- Schumacker, R. E. (2002). Latent Variable Interaction Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(1), 40–54. http://doi.org/10.1207/S15328007SEM0901_3
- Schumacker, R. E., & Rigdon, E. (1995). Testing Interaction Effects in Structural Equation Models. Retrieved from <http://eric.ed.gov/?id=ED384618>
- Scott, D. (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology*, 1(2), 233–247. [http://doi.org/10.1016/0022-2496\(64\)90002-1](http://doi.org/10.1016/0022-2496(64)90002-1)
- Searle, J. R. (1996). *The Construction of Social Reality* (New Ed). Penguin.

- Searle, J. R. (2000). Le débat » Langage, conscience, rationalité : une philosophie naturelle. *Le Débat*, 109(2), 177–192.
- Shallice, T., & Cooper, R. (2011). *The organisation of mind* (Oxford University Press). Oxford.
- Sherry, D. (2011). Thermoscopes, thermometers, and the foundations of measurement. *Studies in History and Philosophy of Science Part A*, 42(4), 509–524.
<http://doi.org/10.1016/j.shpsa.2011.07.001>
- Shimamura, A. P. (2010). Bridging Psychological and Biological Science: The Good, Bad, and Ugly. *Perspectives on Psychological Science*, 5(6), 772–775.
<http://doi.org/10.1177/1745691610388781>
- Sijtsma, K. (2009a). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107–120. <http://doi.org/10.1007/s11336-008-9101-0>
- Sijtsma, K. (2009b). Reliability Beyond Theory and Into Practice. *Psychometrika*, 74(1), 169–173.
<http://doi.org/10.1007/s11336-008-9103-y>
- Sijtsma, K. (2011). Introduction to the measurement of psychological attributes. *Measurement*, 44(7), 1209–1219. <http://doi.org/10.1016/j.measurement.2011.03.019>
- Sijtsma, K. (2012a). Future of Psychometrics: Ask What Psychometrics Can Do for Psychology. *Psychometrika*, 77(1), 4–20. <http://doi.org/10.1007/s11336-011-9242-4>
- Sijtsma, K. (2012b). Psychological measurement between physics and statistics. *Theory & Psychology*, 22(6), 786–809. <http://doi.org/10.1177/0959354312454353>
- Sijtsma, K., & Emons, W. H. M. (2013). Separating models, ideas, and data to avoid a paradox: Rejoinder to Humphry. *Theory & Psychology*, 23(6), 786–796.
<http://doi.org/10.1177/0959354313503724>
- Silvia, P. J., & Guido, E. H. (2001). On introspection and self-perception: Does self-focused attention enable accurate self-knowledge? *Review of General Psychology*, 5(3), 241–269.
<http://doi.org/10.1037/1089-2680.5.3.241>

- Sireci, S. G. (2013). Agreeing on Validity Arguments: Validity Arguments. *Journal of Educational Measurement, 50*(1), 99–104. <http://doi.org/10.1111/jedm.12005>
- Sirotin, Y. B., & Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature, 457*(7228), 475–479. <http://doi.org/10.1038/nature07664>
- Slaney, K. L. (2001). On empirical realism and the defining of theoretical terms. *Journal of Theoretical and Philosophical Psychology, 21*(2), 132–152. <http://doi.org/10.1037/h0091202>
- Slaney, K. L. (2012). Laying the cornerstone of construct validity theory: Herbert Feigl's influence on early specifications. *Theory & Psychology, 22*(3), 290–309.
<http://doi.org/10.1177/0959354311400659>
- Slaney, K. L., & Racine, T. P. (2011). On the ambiguity of concept use in commentaries. *Journal of Theoretical and Philosophical Psychology, 31*(2), 115–125. <http://doi.org/10.1037/a0023427>
- Slaney, K. L., Storey, J. E., & Barnes, J. (2011). Is My Test Valid? Guidelines for the Practicing Psychologist for Evaluating the Psychometric Properties of Measures. *International Journal of Forensic Mental Health, 10*(4), 261–283. <http://doi.org/10.1080/14999013.2011.627086>
- Slaney, & Racine. (2013a). Constructing an understanding of constructs -. *New Ideas in Psychology, 31*(1), 1–3.
- Slaney, & Racine. (2013b). What's in a name? Psychology's ever evasive construct. *New Ideas in Psychology, 31*(1), 4–12.
- Soler, L. (2009). *Introduction à l'épistémologie* (Ellipses). Paris.
- Sosa, E. (1993). Putnam's Pragmatic Realism. *The Journal of Philosophy, 90*(12), 605.
<http://doi.org/10.2307/2940814>
- Soubeyran, A. (2012). *Émergence et problème corps-esprit*. Paris -Sorbonne.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Special Issue on Structural Equation Modeling, 42*(5), 893–898.
<http://doi.org/10.1016/j.paid.2006.09.017>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.

- Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (2006). *Foundations of Measurement Volume III: Representation, Axiomatization, and Invariance*. Dover Publications.
- Tambolo, L. (2015). A tale of three theories: Feyerabend and Popper on progress and the aim of science. *Studies in History and Philosophy of Science Part A*, 51, 33–41.
<http://doi.org/10.1016/j.shpsa.2015.02.005>
- Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 640–657. <http://doi.org/10.1080/10705511.2013.824781>
- Temme, D., Diamantopoulos, A., & Pefegfeidel, V. (2014). Specifying formatively-measured constructs in endogenous positions in structural equation models: Caveats and guidelines for researchers. *International Journal of Research in Marketing*, 31(3), 309–316.
<http://doi.org/10.1016/j.ijresmar.2014.03.002>
- Tenenhaus, M., & Hanafi, M. (2010). A bridge between PLS Path Modeling and multi-block data analysis. In *Handbook of Partial Least Squares* (Springer, pp. 99–124). Londres: V. Esposito Vinzi, W. W. Chin, J. Henseler, H. Wang.
- Tenenhaus, M., Vinzi V.E., Chatelin Y.-M., & Lauro C. (2005). PLS Path Modeling. *Computational Statistics and Data Analysis*, 48, 159–205.
- Tensaout, M., & Guyon, H. (2013). Spécification et estimation d'un mes avec des variables latentes formatives endogènes. Presented at the Congrès SFDS, Toulouse.
- Tensaout, M., Guyon, H., & Hanafi, M. (2011). Estimation des modèles de mesure formatifs avec Path-PLS et l'interprétation confondante. Presented at the Congrès SFDS, Tunis.
- Thiébaud, E. (2015). Variabilité contre différence : à la recherche du sens. In *Différences et variabilités en psychologie*. P.U.R.
- Tiberghien, G. (2007). Entre neurosciences et neurophilosophie : la psychologie cognitive et les sciences cognitives. *Psychologie Française*, 52(3), 279–297.
<http://doi.org/10.1016/j.psfr.2007.05.001>

- Tiercelin, C. (2002). *Hilary Putnam, l'héritage pragmatiste*. Paris: Presses Universitaires de France - PUF.
- Trendler, G. (2009). Measurement Theory, Psychology and the Revolution That Cannot Happen. *Theory & Psychology, 19*(5), 579–599. <http://doi.org/10.1177/0959354309341926>
- Trendler, G. (2013). Measurement in psychology: A case of ignoramus et ignorabimus? A rejoinder. *Theory & Psychology, 23*(5), 591–615. <http://doi.org/10.1177/0959354313490451>
- Ulises Moulines. (2006). *La philosophie des sciences, l'invention d'une discipline* (Rue d'Ulm). Paris: Ecole Normale Supérieure.
- Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842–861.
- Van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review, 118*(2), 339–356. <http://doi.org/10.1037/a0022749>
- Van de Vijver, G. (1997). Emergence et explication. *Intellectica, 25*(2), 7–23.
- Van Orden, G. C., & Kloos, H. (2003). The Module Mistake. *Cortex, 39*(1), 164–166. [http://doi.org/10.1016/S0010-9452\(08\)70092-3](http://doi.org/10.1016/S0010-9452(08)70092-3)
- Van Orden, G. C., Kloos, H., & Wallot, S. (2011). Living in the Pink: Intentionality, Wellbeing, and Complexity. In C. Hooker (Ed.), *Philosophy of Complex Systems* (Vol. 10, pp. 629–672). Amsterdam: North-Holland. Retrieved from <http://www.sciencedirect.com/science/article/pii/B9780444520760500225>
- Van Orden, G. C., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science, 64*(4), S85.
- Varela, F. (1998). Le cerveau n'est pas un ordinateur. *La Recherche, 308*, 109.

- Varela, F., Thompson, E., Rosch, E., & Véronique. (1999). *L'inscription corporelle de l'esprit : Sciences cognitives et expérience humaine*. Paris: Seuil.
- Vautier, S. (2011). The operationalization of general hypotheses versus the discovery of empirical laws in Psychology. *Philosophia Scientiae*, (15-2), 105–122.
<http://doi.org/10.4000/philosophiascientiae.656>
- Vautier, S. (2012). Propos sur la responsabilité scientifique du psychologue : essai d'épistémologie appliquée. *Pratiques Psychologiques*, 18(4), 373–383.
<http://doi.org/10.1016/j.prps.2012.07.001>
- Vautier, S. (2013). How to state general qualitative facts in psychology? *Quality & Quantity*, 47(1), 49–56. <http://doi.org/10.1007/s11135-011-9502-5>
- Vautier, S. (2015a). 3a. Mesurer avec des réponses à des tests : définition. Retrieved September 22, 2015, from <http://epistemo.hypotheses.org/category/projet-de-precis>
- Vautier, S. (2015b). 17a. Combien font deux oranges plus trois citrons ? Le méli-mélo conceptuel de Laveault et Grégoire. Retrieved September 22, 2015, from <http://epistemo.hypotheses.org/category/projet-de-precis>
- Vautier, S. (2015c). 26 a. Le fétichisme des scores psychotechniques. Retrieved from <http://epistemo.hypotheses.org/category/projet-de-precis>
- Vautier, S. (2015d). 35a. Observations objectives vs. subjectives. Retrieved from <http://epistemo.hypotheses.org/category/projet-de-precis/page/3>
- Vautier, S. (2015e). 36a. Comment lire une échelle analogique de douleur ? Retrieved September 22, 2015, from <http://epistemo.hypotheses.org/category/projet-de-precis>
- Vautier, S. (2015f). La psychotechnique des aptitudes. Pour différencier une sociotechnique de l'évaluation sans mesurage et une psychologie balbutiante de la compréhension de la performance. *Pratiques Psychologiques*, 21(1), 1–18.
<http://doi.org/10.1016/j.prps.2015.01.005>

- Vautier, S. (2015g). Une critique du conformisme méthodologique en psychologie générale : on peut aussi chercher des relations générales restrictives. Retrieved from <https://hal.archives-ouvertes.fr/hal-00960258/document>
- Vautier, S., & Afzali, M. H. (2015). Assessing Mood Change with Visual Analogue Scales: Composite vs. Vectorial Approaches.
- Vautier, S., Lacot, É., & Veldhuis, M. (2014). Puzzle-solving in psychology: The neo-Galtonian vs. nomothetic research focuses. *New Ideas in Psychology*, 33, 46–53.
<http://doi.org/10.1016/j.newideapsych.2013.10.002>
- Vautier, S., Veldhuis, M., Lacot, É., & Matton, N. (2012). The ambiguous utility of psychometrics for the interpretative foundation of socially relevant avatars. *Theory & Psychology*, 22(6), 810–822. <http://doi.org/10.1177/0959354312450093>
- Vindras, P., Desmurget, M., & Baraduc, P. (2012). When One Size Does Not Fit All: A Simple Statistical Method to Deal with Across-Individual Variations of Effects. *PLoS ONE*, 7(6), 1–15.
<http://doi.org/10.1371/journal.pone.0039059>
- Vinzi, V. E., Chin, W. W., Henseler, & Wang. (2010). *Handbook of partial least squares*.
- Volken, H. (2007). Les fondements : mesure pour mesure. *Revue européenne des sciences sociales. European Journal of Social Sciences*, (XLV-138), 55–65. <http://doi.org/10.4000/ress.192>
- Vul, E. (2011). Reductionism and Practicality. *Psychological Inquiry*, 22(2), 137–138.
<http://doi.org/10.1080/1047840X.2011.551323>
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3), 274–290. <http://doi.org/10.1111/j.1745-6924.2009.01125.x>
- Vul, E., & Kanwisher, N. (2010). Begging the question: the non-independence error in fMRI data analysis. In *Foundations and Philosophy for Neuroimaging*.

- Wagenmakers, E.-J., & Waldorp, L. (2006). Editors' introduction. *Special Issue on Model Selection: Theoretical Developments and Applications Special Issue on Model Selection: Theoretical Developments and Applications*, 50(2), 99–100. <http://doi.org/10.1016/j.jmp.2005.01.005>
- Wallot, S., & Van Orden, G. (2011). Toward a Lifespan Metric of Reading Fluency. *International Journal of Bifurcation & Chaos in Applied Sciences & Engineering*, 21(4), 1173–1192.
- Wetzels, M., Odekerken-Schröder, G., & van Oppen, C. (2009). Using Pls Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration. *MIS Quarterly*, 33(1), 177–195.
- Widaman, K. F. (2014). Much Ado About Nothing—Or at Best, Very Little. *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 165–168.
<http://doi.org/10.1080/15366367.2014.980109>
- Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia*, 62(2), 100–119. <http://doi.org/10.1007/s11881-012-0067-3>
- Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, 61(12), 1219–1228. <http://doi.org/10.1016/j.jbusres.2008.01.010>
- Willoughby, M. T. (2014). Formative Versus Reflective Measurement of Executive Function Tasks: Response to Commentaries and Another Perspective. *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 173–178. <http://doi.org/10.1080/15366367.2014.981074>
- Willoughby, M. T., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive Functions: Formative Versus Reflective Measurement. *Measurement: Interdisciplinary Research and Perspectives*, 12(3), 69–95. <http://doi.org/10.1080/15366367.2014.929453>
- Wilson, M. (2013). Using the concept of a measurement system to characterize measurement models used in psychometrics. *Measurement*, 46(9), 3766–3774.
<http://doi.org/10.1016/j.measurement.2013.04.005>

- Wilson, M., Mari, L., Maul, A., & Iribarra, D. T. (2015). A comparison of measurement concepts across physical science and social science domains: instrument design, calibration, and measurement. *Journal of Physics: Conference Series*, 588, 012034.
<http://doi.org/10.1088/1742-6596/588/1/012034>
- Wu, W., & Lang, K. M. (2015). Proportionality Assumption in Latent Basis Curve Models: A Cautionary Note. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–15.
<http://doi.org/10.1080/10705511.2014.938578>
- Ziegler, M., & Vautier, S. (2014). A farewell, a welcome, and an unusual exchange. *European Journal of Psychological Assessment*, 30(2), 81–85. <http://doi.org/10.1027/1015-5759/a000203>
- Ziv, Y., & Hotam, Y. (2015). Theory and measure in the psychological field: The case of attachment theory and the strange situation procedure. *Theory & Psychology*, 25(3), 274–291.
<http://doi.org/10.1177/0959354315577970>

Annexes 1 : Code R des simulations

Je ne mets pas tous les codes utilisés, je n'illustre le code générique utilisé à partir du code ci-dessous qui développe le test LR pour le modèle 1 (avec deux réflectives latentes). Le code aurait pu être plus généreux en lignes, notamment sur le codage de paramètres pour les différents modèles comparés ; mais sa construction permet à un lecteur non habitué à R de lire le code plus aisément que si j'avais utilisé des subtilités propres à R (que je ne maîtrise que partiellement).

```

library(mvtnorm)
library(sem)

TEST=matrix(data=NA,nrow=120000,ncol=1)

## Modèle du SEM MICMIC avec SEM
model <- specifyModel()
X1->L1,lamX1,NA
X2->L1,lamX2,NA
X3->L1,lamX3,NA
X4->L1,lamX4,NA
L2->Y1,lamy1,NA
L2->Y2,lamy2,NA
L2->Y3,lamy3,NA
L2->Y4,lamy4,NA
L3->Y5,lamy5,NA
L3->Y6,lamy6,NA
L3->Y7,lamy7,NA
L3->Y8,lamy8,NA
X1<->X1,ERX1,NA
X2<->X2,ERX2,NA
X3<->X3,ERX3,NA
X4<->X4,ERX4,NA
X1<->X2,covX1X2,NA
X1<->X3,covX1X3,NA
X1<->X4,covX1X4,NA
X2<->X3,covX2X3,NA
X2<->X4,covX2X4,NA
X3<->X4,covX3X4,NA
Y1<-> Y1,ERY1,NA
Y2<-> Y2,ERY2,NA
Y3<-> Y3,ERY3,NA
Y4<-> Y4,ERY4,NA
Y5<-> Y5,ERY5,NA
Y6<-> Y6,ERY6,NA
Y7<-> Y7,ERY7,NA
Y8<-> Y8,ERY8,NA
L1 <-> L1,NA,1
L2<->L2,NA,1
L3<->L3,NA,1
L1->L2,C2,NA
L1->L3,C3,NA

#

# Modèle du SEM MICMIC avec SEM
model12 <- specifyModel()
X1->L1,lamX1,NA

```

```

X2->L1,lamX1,NA
X3->L1,lamX1,NA
X4->L1,lamX1,NA
L2->Y1,lamy1,NA
L2->Y2,lamy2,NA
L2->Y3,lamy3,NA
L2->Y4,lamy4,NA
L3->Y5,lamy5,NA
L3->Y6,lamy6,NA
L3->Y7,lamy7,NA
L3->Y8,lamy8,NA
X1<->X1,ERX1,NA
X2<->X2,ERX2,NA
X3<->X3,ERX3,NA
X4<->X4,ERX4,NA
X1<->X2,covX1X2,NA
X1<->X3,covX1X3,NA
X1<->X4,covX1X4,NA
X2<->X3,covX2X3,NA
X2<->X4,covX2X4,NA
X3<->X4,covX3X4,NA
Y1<-> Y1,ERY1,NA
Y2<-> Y2,ERY2,NA
Y3<-> Y3,ERY3,NA
Y4<-> Y4,ERY4,NA
Y5<-> Y5,ERY5,NA
Y6<-> Y6,ERY6,NA
Y7<-> Y7,ERY7,NA
Y8<-> Y8,ERY8,NA
L1 <-> L1,NA,1
L2<->L2,NA,1
L3<->L3,NA,1
L1->L2,C2,NA
L1->L3,C3,NA

```

```
# Modèle du SEM MICMIC avec SEM
```

```
model10 <- specifyModel()
```

```

X1->L1,NA,0
X2->L1,lamX2,NA
X3->L1,lamX3,NA
X4->L1,lamX4,NA
L2->Y1,lamy1,NA
L2->Y2,lamy2,NA
L2->Y3,lamy3,NA
L2->Y4,lamy4,NA
L3->Y5,lamy5,NA
L3->Y6,lamy6,NA
L3->Y7,lamy7,NA
L3->Y8,lamy8,NA
X1<->X1,ERX1,NA
X2<->X2,ERX2,NA
X3<->X3,ERX3,NA
X4<->X4,ERX4,NA
X1<->X2,covX1X2,NA
X1<->X3,covX1X3,NA
X1<->X4,covX1X4,NA
X2<->X3,covX2X3,NA
X2<->X4,covX2X4,NA
X3<->X4,covX3X4,NA

```

```

Y1<-> Y1,ERY1,NA
Y2<-> Y2,ERY2,NA
Y3<-> Y3,ERY3,NA
Y4<-> Y4,ERY4,NA
Y5<-> Y5,ERY5,NA
Y6<-> Y6,ERY6,NA
Y7<-> Y7,ERY7,NA
Y8<-> Y8,ERY8,NA
L1 <-> L1,NA,1
L2<->L2,NA,1
L3<->L3,NA,1
L1->L2,C2,NA
L1->L3,C3,NA

model20 <- specifyModel()
X1->L1,lamX1,NA
X2->L1,NA,0
X3->L1,lamX3,NA
X4->L1,lamX4,NA
L2->Y1,lamy1,NA
L2->Y2,lamy2,NA
L2->Y3,lamy3,NA
L2->Y4,lamy4,NA
L3->Y5,lamy5,NA
L3->Y6,lamy6,NA
L3->Y7,lamy7,NA
L3->Y8,lamy8,NA
X1<->X1,ERX1,NA
X2<->X2,ERX2,NA
X3<->X3,ERX3,NA
X4<->X4,ERX4,NA
X1<->X2,covX1X2,NA
X1<->X3,covX1X3,NA
X1<->X4,covX1X4,NA
X2<->X3,covX2X3,NA
X2<->X4,covX2X4,NA
X3<->X4,covX3X4,NA
Y1<-> Y1,ERY1,NA
Y2<-> Y2,ERY2,NA
Y3<-> Y3,ERY3,NA
Y4<-> Y4,ERY4,NA
Y5<-> Y5,ERY5,NA
Y6<-> Y6,ERY6,NA
Y7<-> Y7,ERY7,NA
Y8<-> Y8,ERY8,NA
L1 <-> L1,NA,1
L2<->L2,NA,1
L3<->L3,NA,1
L1->L2,C2,NA
L1->L3,C3,NA

model30 <- specifyModel()
X1->L1,lamX1,NA
X2->L1,lamX2,NA
X3->L1,NA,0
X4->L1,lamX4,NA
L2->Y1,lamy1,NA
L2->Y2,lamy2,NA
L2->Y3,lamy3,NA
L2->Y4,lamy4,NA

```

```

L3->Y5,lamy5,NA
L3->Y6,lamy6,NA
L3->Y7,lamy7,NA
L3->Y8,lamy8,NA
X1<->X1,ERX1,NA
X2<->X2,ERX2,NA
X3<->X3,ERX3,NA
X4<->X4,ERX4,NA
X1<->X2,covX1X2,NA
X1<->X3,covX1X3,NA
X1<->X4,covX1X4,NA
X2<->X3,covX2X3,NA
X2<->X4,covX2X4,NA
X3<->X4,covX3X4,NA
Y1<-> Y1,ERY1,NA
Y2<-> Y2,ERY2,NA
Y3<-> Y3,ERY3,NA
Y4<-> Y4,ERY4,NA
Y5<-> Y5,ERY5,NA
Y6<-> Y6,ERY6,NA
Y7<-> Y7,ERY7,NA
Y8<-> Y8,ERY8,NA
L1 <-> L1,NA,1
L2<->L2,NA,1
L3<->L3,NA,1
L1->L2,C2,NA
L1->L3,C3,NA

model40 <- specifyModel()
X1->L1,lamX1,NA
X2->L1,lamX2,NA
X3->L1,lamX3,NA
X4->L1,NA,0
L2->Y1,lamy1,NA
L2->Y2,lamy2,NA
L2->Y3,lamy3,NA
L2->Y4,lamy4,NA
L3->Y5,lamy5,NA
L3->Y6,lamy6,NA
L3->Y7,lamy7,NA
L3->Y8,lamy8,NA
X1<->X1,ERX1,NA
X2<->X2,ERX2,NA
X3<->X3,ERX3,NA
X4<->X4,ERX4,NA
X1<->X2,covX1X2,NA
X1<->X3,covX1X3,NA
X1<->X4,covX1X4,NA
X2<->X3,covX2X3,NA
X2<->X4,covX2X4,NA
X3<->X4,covX3X4,NA
Y1<-> Y1,ERY1,NA
Y2<-> Y2,ERY2,NA
Y3<-> Y3,ERY3,NA
Y4<-> Y4,ERY4,NA
Y5<-> Y5,ERY5,NA
Y6<-> Y6,ERY6,NA
Y7<-> Y7,ERY7,NA
Y8<-> Y8,ERY8,NA
L1 <-> L1,NA,1

```

```

L2<->L2,NA,1
L3<->L3,NA,1
L1->L2,C2,NA
L1->L3,C3,NA

resultat=matrix(data = NA, nrow = 72000, ncol = 7, byrow = TRUE)
colnames(resultat)=c("taille","cov","R sq","Cor1", "Cor2","simu","LR")

## NOMBRE DE SIMULATIONS
NSIMU=500

##Compteur initialisé

K1=0
K2=0

## COVARIANCE entre les Indicateurs Formatifs
cov=0.0

SIGMA=matrix(c(rep(c(1,rep(cov,4)),3),1),ncol=4)

## TAILLE ECHANTILLON

for(taille in c(1000,500,250)) {

### CORR0=R2 de la Formative par rapport à ses indicateurs

  for(iRSQ in 1:6) {
    CORR0=0.4 + (iRSQ-1)*0.10

    ## Corrélations des manifestes réflectives de la latente formative

      for(icor1 in 1:8){

        cor1=0.2+(icor1-1)*0.1

        cor2=cor1

      }

    ##DEBUT DES SIMMULATIONS

      for(isimu in 1:NSIMU){

        K1=K1+1

        resultat[K1,2]=cov
        resultat[K1,1]=taille
        resultat[K1,3]=CORR0
        resultat[K1,4]=cor1
        resultat[K1,5]=cor2
        resultat[K1,6]=isimu

      }

    ##La matrice des manifestes formatives :
      FORMAT=data.frame(scale(rmvnorm(taille,c(0,0,0,0),SIGMA)))

    ##CREATION DE LA LATENTE avec un R2=CORR0 par rapport aux manifestes formatives
      LAT=scale(scale(rowSums(FORMAT))+rnorm(taille,0,((1-
CORR0)/CORR0)^(1/2)))

```



```

#### LATENTES REFLECTIVES #####

LAT1=scale(LAT+rnorm(taille,0,((1-cor1^2)/cor1^2)^(1/2)))
LAT2=scale(LAT+rnorm(taille,0,((1-cor2^2)/cor2^2)^(1/2)))

### MANIFESTES des LATENTES REFLECTIVES
#####
Y1=scale(LAT1+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y2=scale(LAT1+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y3=scale(LAT1+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y4=scale(LAT1+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y5=scale(LAT2+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y6=scale(LAT2+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y7=scale(LAT2+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))
Y8=scale(LAT2+rnorm(taille,0,((1-0.9^2)/0.9^2)^(1/2)))

MANIF=as.data.frame(cbind(FORMAT,Y1,Y2,Y3,Y4,Y5,Y6,Y7,Y8))

colnames(MANIF)=c("X1","X2","X3","X4","Y1","Y2","Y3","Y4","Y5","Y6","Y7","Y8")
COR=cov(as.matrix(MANIF))

#####
#####
## Modèle CB-SEM

RSLT=sem(model=model,COR,N=taille,standardized=TRUE,maxiter=100)

RSLT12=sem(model=model12,COR,N=taille,standardized=TRUE,maxiter=100)

RSLT10=sem(model=model10,COR,N=taille,standardized=TRUE,maxiter=100)

RSLT20=sem(model=model20,COR,N=taille,standardized=TRUE,maxiter=100)

RSLT30=sem(model=model30,COR,N=taille,standardized=TRUE,maxiter=100)

RSLT40=sem(model=model40,COR,N=taille,standardized=TRUE,maxiter=100)

CONV1=0
CONV12=0

LRT1=1

if(det(RSLT$vcov)>0 & RSLT$convergence & !is.na(RSLT$vcov)[1,1]){
  CONV1=1

  if(det(RSLT12$vcov)>0 & RSLT12$convergence & !is.na(RSLT12$vcov)[1,1]){
    CONV12=1
    if(summary(RSLT12)$chisq -summary(RSLT)$chisq > 11.34) LRT1=0
  }

  if(det(RSLT10$vcov)>0 & RSLT10$convergence & !is.na(RSLT10$vcov)[1,1]){
    if(summary(RSLT10)$chisq -summary(RSLT)$chisq <=6.63) LRT1=0
  }

  if(det(RSLT20$vcov)>0 & RSLT20$convergence & !is.na(RSLT20$vcov)[1,1]){
    if(summary(RSLT20)$chisq -summary(RSLT)$chisq <=6.63) LRT1=0
  }
}

```

```

}

if(det(RSLT30$vcov)>0 & RSLT30$convergence & !is.na(RSLT30$vcov)[1,1]){
  if(summary(RSLT30)$chisq -summary(RSLT)$chisq <=6.63) LRT1=0
}

if(det(RSLT40$vcov)>0 & RSLT40$convergence & !is.na(RSLT40$vcov)[1,1]){
  if(summary(RSLT40)$chisq -summary(RSLT)$chisq <=6.63) LRT1=0
}

}

if(CONV1==0) LRT1=0
if(CONV12==0) LRT1=0

resultat[K1,7]=LRT1

print(K1)
}
}
}
}
write.csv2(resultat,file="resultats reflective 2 reflectives LR CHI2 3 dl.csv", row.names=FALSE)

```


Résumé : Ma thèse développe deux parties. La première considère que la psychologie expérimentale doit clarifier son positionnement épistémologique pour clarifier la validation formelle de sa démarche, sans forcément devoir se référer au cadre de la Science Physique. A partir d'une réflexion critique, je propose de décaler le cadre épistémologique en psychologie et de poser clairement un cadre pragmatique-réaliste. La thèse essentielle défendue dans ce travail est : 1/ les propriétés mentales doivent être comprises comme des phénomènes émergents, ce qui implique que leurs analyses ne peuvent se faire ni au niveau neuronal, ni au niveau de la dynamique interne de processus cognitifs, mais nécessairement au niveau de ces phénomènes émergents ; 2/ pour analyser les propriétés mentales comme formes émergentes, la psychométrie a besoin d'utiliser de concepts qui sont en tension permanente entre une objectivité et une intersubjectivité ; en conséquence, la psychométrie doit affirmer une démarche pragmatiste-réaliste, en rupture avec l'empirisme-réaliste classique ; 3/ une approche pragmatiste-réaliste, basée entre autre sur l'abduction, permet de dépasser les contradictions pointées dans la littérature académique sur les propriétés mentales et leurs mesures ; 4/ un cadre de mesure de propriétés mentales par des variables latentes devient dès lors possible si ce cadre est compris lui aussi comme pragmatiste-réaliste ; 5/ mais ce recours au pragmatisme-réaliste renvoie en conséquence une critique à la fois des modèles avec variables latentes développés dans la littérature académique et les usages sociaux de ces modèles.

La seconde partie de ma thèse porte sur un cadre particulier de formalisation des variables latentes : le cadre formatif. Je développe des simulations Monte Carlo pour vérifier le spectre des paramètres permettant une mesure formative efficiente dans le cadre d'un positionnement réaliste-empirique.

Title: Mental processes and latent variables: an epistemological and methodological reflection

Abstract: My thesis considers that experimental psychology must clarify its epistemological position to clarify the formal validation of its approach, without necessarily having to refer to the framework of Science Physics. From a critical reflection, I propose to shift the epistemological framework in psychology and clearly pose a pragmatic-realistic framework. The main thesis of this work is: 1 / mental properties must be understood as emerging phenomena, which implies that their analysis can not be done nor at the neuronal level, nor at the internal dynamics of cognitive processes, but necessarily at these emerging phenomena; 2 / to analyze the mental properties as emerging forms, psychometrics need to use concepts that are in permanent tension between objectivity and intersubjectivity; accordingly, psychometrics must assert a pragmatic-realist approach, breaking with classical empiricism-realistic; 3 / a pragmatist-realistic approach, based among other things on the abduction, can overcome the contradictions pointed in the academic literature on mental properties and their measurements; 4 / a framework for measuring mental properties by latent variables becomes possible if the framework is also understood as a pragmatic-realist; 5 / but use realistic-pragmatic returns accordingly critical of both models with latent variables developed in the academic literature and the social uses of these models. The second part of my thesis focuses on a specific part of formalization of latent variables: the formative model. I develop Monte Carlo simulations to check the range of parameters for efficient formative measure as part of a realistic-empirical positioning.

Mots clés (français) : Emergence, Epistémologie, Modèles de mesures formatives, Pragmatisme, Processus mentaux, Simulations Monte Carlo, Variables latentes

Keywords: Emergence, Epistemology, Formative measurement models, Latent variable, Mental processes, Monte Carlo simulations, Pragmatism