



HAL
open science

Le rôle des émotions morales dans l'analyse des dilemmes sociaux : la gratitude au coeur des relations entre agents économiques

Romain Jourdheuil

► **To cite this version:**

Romain Jourdheuil. Le rôle des émotions morales dans l'analyse des dilemmes sociaux : la gratitude au coeur des relations entre agents économiques. Economies et finances. Université de Bordeaux, 2016. Français. NNT : 2016BORD0296 . tel-01524693

HAL Id: tel-01524693

<https://theses.hal.science/tel-01524693>

Submitted on 18 May 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.

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE

DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE ENTREPRISE, ÉCONOMIE, SOCIÉTÉ (ED N°42)

SPÉCIALITÉ SCIENCES ÉCONOMIQUES

Par Romain Jourdheuil

**LE RÔLE DES ÉMOTIONS MORALES DANS L'ANALYSE DES
DILEMMES SOCIAUX**

La Gratitude au Cœur des Relations entre Agents Économiques

Sous la direction de : **Emmanuel PETIT**

Soutenue le 7 Décembre 2016

Membres du jury :

Mme AUBERT, Cécile

Professeur des Universités, Université de Bordeaux, **Examinatrice**

M. JACQUEMET, Nicolas

Professeur des Universités, Université Paris 1, **Rapporteur**

M. MASCRET, David

Professeur des Universités, Université de Rennes 1, **Rapporteur**

M. PETIT, Emmanuel

Professeur des Universités, Université de Bordeaux, **Directeur de thèse**

M. ROUILLON, Sébastien

Professeur des Universités, Université de Bordeaux, **Examineur**

Mme TCHERKASSOF, Anna

Maître de Conférences, Université Grenoble Alpes, **Examinatrice**

“The University is not to provide any approval or disapproval regarding the opinions this PhD dissertation includes. These opinions must be considered as being solely those of their authors.”

« L'Université n'entend ni approuver, ni désapprouver les opinions particulières émises dans cette thèse. Ces opinions sont considérées comme propres à leur auteur. »

Titre : LE RÔLE DES ÉMOTIONS MORALES DANS L'ANALYSE DES DILEMMES SOCIAUX

Résumé : Cette thèse étudie le rôle des émotions morales dans le choix des agents lorsqu'ils font face à un dilemme social. Lors des deux premiers chapitres, nous présentons une revue de littérature étendue, qui mobilise des recherches en économie expérimentale, psychologie sociale et gestion des organisations, afin de cerner comment les émotions morales, et en particulier la gratitude, influent sur la prise de décision des agents économiques. Dans le troisième chapitre, nous nous intéressons, à travers une revue des travaux en économie comportementale, à l'intégration progressive du rôle des émotions dans les modèles économiques. Nous montrons ainsi en quoi cette intégration, concomitante à l'apparition des modèles dits de « préférences sociales », permet de représenter de manière plus réaliste le processus de décision des agents, dans des situations où les comportements économiques observés expérimentalement remettent en cause une vision égoïste et matérialiste des êtres humains. Dans le quatrième chapitre, nous présentons en détail le jeu du bien public, ainsi que les principaux résultats expérimentaux qui lui sont associés, pour mieux saisir en quoi il constitue un cadre propice à l'analyse formelle de l'influence des émotions morales. Enfin, notre travail aboutit, dans les cinquième et sixième chapitres, à l'élaboration d'un modèle original en économie comportementale, qui vise à analyser comment l'émergence des émotions morales peut influencer, à la fois en statique et en dynamique, le choix stratégique des individus dans le jeu du bien public et comment l'ajout d'une phase de récompense peut renforcer la coopération entre les partenaires.

Mots clés : Émotions morales, Gratitude, Dilemme social, Jeu du bien public, Récompense, Dynamique

Title: THE ROLE OF MORAL EMOTIONS IN THE ANALYSIS OF SOCIAL DILEMMAS

Abstract: This thesis studies the role of moral emotions in the decision making of agents when they face a social dilemma. The first two chapters present a large literature review, which gathers research in experimental economics, in social psychology and in the management of organizations, in order to delineate the way moral emotions, and specifically gratitude, are able to influence the economical decision process. In the third chapter, through a review of works in the field of behavioral economics, we take an interest in the progressive integration of moral emotions into economic modeling. We thus show how this integration, which is concurrent with the emergence of social preferences models, allows agents' decision-making processes to be represented in a more realistic way, in situations where the behaviors that are observed empirically challenge an egoistic and materialistic vision of people. In chapter four, the public good game is presented in detail, along with the main experimental results associated with it, in order to understand why this game suits the formal analysis of moral emotions' influence. Finally, our work culminates in the development of an original model in behavioral economics, during chapters five and six, which aims at assessing how the emergence of moral emotions can influence, both statically and dynamically, the strategic choice of individuals in the public good game, and how the introduction of a reward phase can promote cooperation and help individuals to develop good interpersonal relationships.

Keywords: Moral emotions, Gratitude, Social dilemma, Public good game, Reward, Dynamic

GREThA UMR CNRS 5113 - Université de Bordeaux

Avenue Léon Duguit, 33608 Pessac Cedex, France

Remerciements

Tout d'abord, je souhaite exprimer ma profonde gratitude à l'égard de mon directeur de thèse, Emmanuel Petit, pour avoir cru en moi depuis le départ, pour toute l'énergie qu'il m'a consacrée, pour l'aide et les conseils qu'il a su m'apporter, ainsi que pour avoir su trouver la recette pour que notre collaboration, durant ces cinq années, me permette d'exprimer au mieux mon potentiel. Je tiens également à remercier Nicolas Jacquemet et David Masclat pour l'intérêt qu'ils ont porté à mon travail en acceptant d'en être les rapporteurs. Mes remerciements vont aussi à Cécile Aubert, Sébastien Rouillon et Anna Tcherkassof qui me font l'honneur de prendre part au jury de de cette thèse.

Je souhaite également remercier l'Université de Bordeaux, l'école doctorale EES n°42 et le laboratoire du GREThA pour m'avoir recruté, financé et offert un tel cadre de travail pour mener à bien ce projet. Je remercie ainsi l'ensemble des membres du GREThA et notamment ses personnels administratifs, au premier rang desquels se trouve son directeur Marc-Alexandre Sénégas.

Pendant mon parcours doctoral à l'Université de Bordeaux, j'ai également eu l'occasion d'intégrer plusieurs équipes pédagogiques formidables et ainsi de goûter aux plaisirs de l'enseignement. Je tiens donc à remercier Olivier Baron, Hervé Hocquard et Jean-Christophe Pereau pour m'avoir fait confiance et pour m'avoir donné l'opportunité d'enseigner des matières qui me passionnent. J'ai eu beaucoup de plaisir à travailler avec eux et avec les autres membres de ces équipes.

Je souhaite également exprimer ma gratitude à ceux qui ont partagé mon quotidien tout au long de cet étrange voyage. En premier lieu, je remercie mes collègues du bureau D315, Guillaume Pastureau, qui m'a montré la voie, François Chounet, pour sa philosophie, Hamza Abadi et Robin Vos, mes collègues les plus jeunes, pour le vent de fraîcheur qu'ils ont apporté à ce bureau, Dan Tran, pour les bons moments qu'on a partagés et pour l'aide qu'il m'a apportée et, enfin, Huihui Ding, avec qui j'ai vécu cette expérience (presque) de bout en bout. Bien sûr, j'ai partagé et échangé avec d'autres collègues doctorants et, parmi ceux-là, je tiens particulièrement à remercier Alexandre Berthe, pour nos discussions animées, François Maublanc, pour nos échanges et nos prémices de collaboration, Luc Elie, pour son humour et pour nos échanges musicaux, Marianne Lanöe, Pauline Lectard et Tsiry Andrianampiarivo, avec qui j'ai débuté cette aventure, Samuel Klebaner, pour ses convictions d'enseignant et, enfin, Viola Lamani, Natasha et Pierre Fauvet, Rémi Delille, Laurent Bergé, Emmanuel Lorenzon, Pierre Levasseur et Nicolas Yol pour leur sympathie.

Un grand merci à tous mes coéquipiers et à tous mes coachs, au basket comme au volley. Je vous dois énormément et cela a été un plaisir d'avoir pu me ressourcer et évacuer auprès de vous. Vous m'avez permis d'observer concrètement ce que cela signifie d'être dans une équipe, d'avoir un objectif commun, de coopérer et de se battre pour y parvenir. Plus encore, j'ai trouvé auprès de vous de la fraternité et de la camaraderie. On est ensemble.

Je tiens également à remercier mes proches et mes amis les plus chers. En particulier, je voudrais dire merci à Maud Cantin, Anaïs Castel, Oriane Guy, Stéphanie Laclau, Killian Bédis, Etienne Bourdin, Eric Dantas, Bastien Grange, Thomas Salvador et Martin Véron pour avoir été une présence salvatrice à mes côtés pendant toutes ces années. Un remerciement spécial à Julie Pacaud, pour son attention bienveillante et son soutien psychologique (j'aurai finalement tenu mon pari) et à Morgane Sifre pour avoir su me convaincre de ne jamais abandonner, même dans les pires moments de doute. Je souhaite également remercier André Leroux pour avoir contribué à ce que je me lance dans cette voie, ainsi que Christine et Manon Leroux pour les bons moments qu'on a passés tous ensemble. Merci également à Gigi Foster pour son expérience et pour tout ce qu'elle a pu me transmettre.

Pour finir, je voudrais remercier ma famille. Surtout, je souhaite dire merci à mes frères, Julien et Valentin (et Polo), pour leur humour, pour leur dérision à mon sujet qui me fait avancer dans l'humilité et pour l'amour qu'on se porte mutuellement. Je remercie mes cousins, Aurélien et Marius Longo, pour leur créativité et pour notre proximité, malgré l'éloignement géographique. Je veux aussi remercier mes grands-parents, Arlette et Serge, qui sont toujours si fiers de moi. Un très grand merci à ma tante Cécile, pour son immense travail de relecture et pour le temps et le soutien qu'elle m'a consacrés. Enfin, je voudrais remercier mes deux parents, pour m'avoir toujours encouragé, soutenu et épaulé dans la voie que j'ai choisie, ainsi que pour tout l'amour qu'ils m'ont prodigué. Merci.

Khanitha, je tiens à te remercier pour m'avoir supporté dans les années les plus difficiles de cette épreuve et, surtout, je voudrais te dire merci pour ton amour et pour le fait de m'avoir rendu meilleur.

Table des matières

Liste des tableaux	v
Liste des figures	vii
I. Introduction générale	1
II. Le rôle des émotions morales dans la prise de décision économique	7
1. Les émotions morales : un système d'outils au service de la décision	9
1.1. Introduction	9
1.2. Les enseignements de la théorie des émotions morales	10
1.2.1. Qu'est-ce qu'une émotion ?	10
1.2.2. La spécificité des émotions morales	13
1.3. Le rôle des émotions morales dans les interactions sociales	30
1.3.1. Le rôle de la sanction morale affective dans l'émergence des comportements sociaux	31
1.3.2. La promotion morale affective des normes sociales ou morales	42
1.4. Conclusion	45
2. La Gratitude et le Côté Radieux des Emotions Morales	47
2.1. Introduction	47
2.2. La gratitude et les relations interpersonnelles	49
2.2.1. La réciprocité positive	49
2.2.2. L'influence de la gratitude sur la qualité des relations entre agents	51
2.2.3. La gratitude comme un moyen de protéger et de préserver les relations	56
2.3. La performance par le bien-être et par la promotion morale affective des comportements coopératifs	59
2.3.1. Gratitude et bien-être	59
2.3.2. Le bien-être est-il décisif dans la performance ?	63
2.3.3. L'influence de la gratitude sur la motivation, un élément clé de la performance et du bien-être	63

III. La Modélisation des Emotions Morales dans le Processus de Décision Economique	71
3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Econo- mique	73
3.1. Introduction	73
3.1.1. L'émergence et le maintien des normes	74
3.2. L'intégration implicite des émotions morales dans les modèles économiques	85
3.2.1. La prise en compte de la norme d'équité	86
3.2.2. Le rôle des intentions et des attentes	91
3.2.3. Le rôle des émotions dans la dynamique des liens sociaux	96
3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique	101
3.3.1. Les modèles centrés sur le rôle d'une émotion particulière	101
3.3.2. Vers des modélisations originales intégrant plusieurs émotions morales	115
3.3.3. Théorie du choix rationnel et prise en compte du rôle des normes et émotions morales	129
3.4. Conclusion	131
4. Le jeu du bien public : un cadre approprié à la représentation du rôle éco- nomique de la gratitude	133
4.1. Le jeu du bien public	133
4.2. Résultats expérimentaux	135
4.2.1. Différentes déclinaisons du jeu	136
4.2.2. La composition dynamique des groupes	137
4.3. L'ajout d'une phase de punition et/ou de récompense	142
4.4. Régulation, communication et leadership	148
4.5. Les atouts du jeu du bien public	151
IV. Le Rôle des Emotions Morales dans l'Analyse Formelle du Jeu du Bien Public avec Phase de Récompense	153
5. Le Cas Statique	155
5.1. Préambule au modèle : Structure du jeu du bien public	155
5.1.1. Structure formelle du jeu	155
5.2. Le Modèle basique à deux joueurs	157
5.2.1. Le jeu du bien public dans sa version standard	158
5.2.2. Le jeu du bien public avec phase de récompense	185
5.3. Généralisation	203
5.3.1. Le jeu du bien public dans sa version standard	203
5.3.2. Le jeu du bien public avec phase de récompense	206
5.4. Conclusion	218

6. L'évolution dynamique des relations entre agents dans le jeu du bien public	221
6.1. Formalisation	223
6.1.1. Hypothèses	223
6.2. Simulations et résultats	231
6.2.1. Résultats pour des profils de joueurs déterminés	232
6.2.2. Résultats pour des profils de joueurs générés aléatoirement	238
6.3. Discussion	241
V. Conclusion Générale	245
Bibliographie	249
VI. Annexe	265

Liste des tableaux

5.1. Les différentes catégories « d'orientation sociale »	164
5.2. Les valeurs prises par δ en fonction de « l'orientation sociale » d'un individu	164
5.3. Exemple de paramétrisation - JBP standard - Cas symétrique	175
5.4. Résultats du JBP standard - Cas symétrique	176
5.5. Exemple de paramétrisation - JBP standard - Cas hétérogène	181
5.6. Résultats du JBP standard - Cas hétérogène	182
5.7. Exemple de paramétrisation - JBP Récompense - Cas symétrique	193
5.8. Résultat du JBP récompense - Cas symétrique	194
5.9. Exemple de paramétrisation - JBP récompense - Cas hétérogène	199
5.10. Résultats du JBP récompense - Cas hétérogène	200
5.11. Exemple de paramétrisation - Modèle généralisé	211
5.12. Résultats - JBP standard - Modèle généralisé	211
5.13. Résultats - JBP récompense - Modèle généralisé	212
5.14. Résultats - JBP standard - Communication	213
5.15. Résultats - JBP récompense - Communication	213
5.16. Résultats - JBP standard - Face à face	214
5.17. Résultats - JBP récompense - Face à face	214
5.18. Modèle généralisé - Simulations sur 100 groupes aléatoirement générés . . .	217
6.1. Exemple de paramétrisation - Modèle dynamique	232

Table des figures

3.3.1. Arbre du jeu de confiance dans le modèle de Tadelis (2008)	106
5.2.1. Fonctions de meilleure réponse	172
5.2.2. Variations de y^* en fonction de β	179
5.2.3. Variations de y^* en fonction de α	180
5.3.1. Fonction de densité de probabilité - Distribution normale $\mu = 0.5$ et $\sigma = 0.125$	215
5.3.2. Fonction de densité de probabilité - Distribution normale asymétrique $\mu =$ 0.5, $\sigma = 0.125$ et $\alpha = -10$	216
5.3.3. Fonction de densité de probabilité - Distribution normale asymétrique $\mu =$ 0.5, $\sigma = 0.125$ et $\alpha = -2$	216
6.2.1. Evolution des niveaux de contribution - JBP standard	232
6.2.2. Evolution du capital de sympathie - JBP standard	233
6.2.3. Evolution des niveaux de contribution - Subvention sur 5 périodes	234
6.2.4. Evolution du capital de sympathie - Subvention sur 5 périodes	235
6.2.5. Evolution des niveaux de contribution - JBP récompense	236
6.2.6. Evolution du capital de sympathie - JBP récompense	237
6.2.7. Evolution des niveaux de récompense versés et reçus	237
6.2.8. Evolution du niveau moyen de contribution - JBP standard	238
6.2.9. Evolution du capital de sympathie moyen - JBP standard	239
6.2.10. Evolution du niveau moyen de contribution - JBP récompense	239
6.2.11. Evolution du niveau moyen de récompense (versé) - JBP récompense	240
6.2.12. Evolution du capital de sympathie moyen - JBP récompense	240

Première partie .
Introduction générale

Introduction générale

« La gratitude est le secret de la vie. L'essentiel est de remercier pour tout. Celui qui a appris cela, sait ce que vivre signifie. Il a pénétré le profond mystère de la vie. »

Albert Schweitzer

Dès lors que l'on se trouve en interaction avec autrui, notre intérêt personnel entre presque mécaniquement en conflit avec l'intérêt de l'autre ou, plus généralement, avec l'intérêt collectif. En effet, même lorsque l'interaction a vocation à être mutuellement bénéfique, il existe presque toujours des moyens de tirer la couverture à soi. Par exemple, lorsque l'on réalise un échange avec un autre individu, la négociation est le processus qui va permettre aux différentes parties de confronter leurs aspirations et de faire valoir leur propre intérêt. De plus, lors de la mise en application de l'échange, même en présence d'un contrat, il subsiste souvent des possibilités, pour l'une des personnes impliquées, d'abuser l'autre partie à des fins personnelles. Par ailleurs, au cours d'une collaboration ou d'un travail d'équipe, il est presque toujours possible de bénéficier de la contribution des autres et il existe donc une incitation économique à tenter d'exploiter la présence de telles externalités. Aussi, par nature, certaines situations exacerbent l'amplitude d'un tel dilemme intérieur. Dans la littérature en économie, ces situations sont appelées les dilemmes sociaux, dont une illustration célèbre est le fameux dilemme du prisonnier, utilisé dans le champ très large des sciences sociales (économie, psychologie, philosophie, sociologie).

La question des dilemmes sociaux soulève, de façon sous-jacente, la question de la capacité des agents économiques à être coopératifs. La coopération est l'un des fondements de l'évolution de l'espèce humaine et, encore aujourd'hui, elle reste une des conditions nécessaires à sa survie. Ainsi, certains défis majeurs que les civilisations humaines vont devoir relever et qui nous concerne tous, nécessitent la mise en place d'un effort coopératif pour être surmontés. Il s'agit par exemple de la question environnementale (recyclage, diminution des gaz à effet de serre etc.), des questions de santé (financement du système de santé, diminution des agents nocifs, comme les pesticides, dans les produits de consommation etc.), ou encore la question de la gestion des migrants. Plus généralement, la coopération entre individus fait partie des conditions *sine qua non* de la réussite de tout projet collaboratif, qu'il soit professionnel ou associatif, ou dès lors que l'on souhaite entretenir de bonnes relations avec ses proches, sa famille ou encore son conjoint.

Toutefois, la nature égoïste qui caractérise toute personne, à laquelle Adam Smith fait référence lorsqu'il évoque, dans la « *richesse des nations* » (1776), « l'amour de soi », peut représenter un obstacle à la mise en œuvre de la coopération au sein des groupes d'agents. Ainsi, en sciences économiques, notamment en théorie des jeux, il a initialement été considéré que, dans les dilemmes sociaux, aucune coopération n'était possible, du fait de la conception même des individus qui y est traditionnellement retenue, à savoir des agents certes rationnels, mais aussi et surtout égoïstes. Pour autant, les êtres humains

sont des êtres sociaux et personne n'est capable de survivre seul, il est donc nécessaire que l'on puisse communiquer les uns avec les autres, de façon pacifique. Pour Adam Smith, cette fois dans une œuvre antérieure, moins connue de la majorité des économistes contemporains, la « *théorie des sentiments moraux* » (1759), ceci est possible grâce à la sympathie, un processus affectif qui peut être assimilé aujourd'hui à une forme d'empathie, via laquelle les individus sont capables de partager les sentiments ressentis par autrui. La sympathie, selon Smith, permet également une certaine proximité entre deux êtres par une reconnaissance simultanée de ce que l'un et l'autre ressentent. Elle correspond, enfin, à un engagement émotionnel réciproque aidant les parties à se coordonner et à se trouver en accord, ce qui lui confère un rôle central dans les rapports sociaux. Cette capacité à partager les émotions d'autrui semble donc être un des fondements de la coopération, à travers le goût que l'on peut trouver à s'aimer soi-même encore davantage, en raison des émotions positives que l'on peut susciter chez autrui par l'adoption de comportements prosociaux.

Plus récemment, l'avènement de l'économie expérimentale et de la psychologie sociale, permettant de tester scientifiquement le comportement des individus dans les dilemmes sociaux, représentés sous la forme de jeux économiques tels que le dilemme du prisonnier, le jeu de l'ultimatum et du dictateur ou encore le jeu du bien public, a permis de mettre en évidence que de les prédictions du modèle classique, en théorie des jeux, s'avéraient globalement erronées (Camerer, 2003). Ces différents résultats ont ainsi conduit certains économistes à repenser la représentation traditionnelle des êtres humains et à s'interroger sur ce qui rend possible la coopération entre les individus dans ces situations particulières. Ainsi, en préservant le cadre canonique de la rationalité, les modèles issus de l'économie comportementale supposent que la motivation égoïste n'est pas la seule qui existe dans le processus de décision des agents. Plus exactement, ces modèles prennent en compte une pluralité de motivations qui peuvent varier en fonction du contexte de la décision et de la personnalité des acteurs. Ainsi, lors d'une interaction sociale, outre le bien-être matériel égoïste, ont été pris en compte, dans la littérature, des motivations additionnelles concernant l'équité et la justice (Fehr et Schmidt, 1999), les considérations de réciprocité entre agents (Rabin, 1993) (en lien avec le rôle des attentes et des intentions), le rapport qu'entretiennent les individus avec la présence de normes sociales et morales dans leurs sociétés (Akerlof, 1980), l'altruisme et le « goût de donner » (Andreoni, 1989) ou encore l'existence d'un lien social entre les agents (van Dijk et van Winden, 1997). Cet élargissement des sources de satisfaction de l'individu, à travers la prise en compte de la diversité de ses motivations, a alors permis à ces différents modèles, de gagner en réalisme et en capacité prédictive, conformément au projet méthodologique de l'économie expérimentale (Guala, 2005).

En filigrane, l'ensemble de ces motivations additionnelles fonctionnent sur la base de la capacité d'empathie (ou sympathie, chez Smith) des agents, mais aussi grâce à l'émergence de certains sentiments et affects qui ont évolués, chez les êtres humains, pour augmenter leur chance de survie, en rendant possible la coopération. En particulier, la famille des émotions morales (Haidt, 2003) comprend les principaux sentiments participant au (bon) déroulement de la vie en société. Parmi ces émotions, on trouve l'indignation (juste colère), la culpabilité, la honte, le dégoût (moral), la fierté, l'admiration ou encore la gratitude. Chacune de ces émotions occupe une fonction particulière dans le processus de décision des agents. Elles se complètent pour permettre aux individus d'adapter leurs

comportements en fonction du contexte spécifique de chaque situation. Il existe ainsi, au sein de la littérature des émotions, dans des disciplines proches de l'économie (psychologie, neurosciences, philosophie), une littérature abondante qui interroge le rôle des ces émotions morales dans l'émergence des comportements prosociaux. Dans l'analyse économique, ce questionnement est aujourd'hui devenu central dans le champ de l'économie du comportement, comme en atteste certains travaux récents sur la culpabilité, la honte ou même la colère.

L'objet de ce travail de thèse est donc d'étudier le rôle spécifique des émotions morales dans le processus de décision des agents économiques, lorsqu'ils sont placés dans des situations de dilemme social. Il s'agit notamment de montrer en quoi ces émotions sont primordiales dans la promotion de la coopération. En outre, malgré l'intérêt précoce porté par Adam Smith à l'émotion à valence positive qu'est la gratitude, pour sa capacité à faciliter les interactions sociales, l'influence de cette émotion a été largement négligée, en économie, que ce soit dans les travaux expérimentaux ou dans les travaux de modélisation. C'est de façon plus générale le domaine des émotions à valence positive qui, comme c'est le cas en psychologie, a été relativement négligé dans la littérature. Convaincu de l'intérêt central que représente l'émotion de gratitude, à la fois pour son utilité dans la promotion des comportements prosociaux, pour son rôle dans la dynamique des relations interindividuelles, mais aussi pour sa capacité à être une source de bien-être pour les agents, nous accordons une attention particulière à cette émotion, tout au long de ce travail de thèse.

Pour répondre à cette problématique, nous présentons, lors des deux premiers chapitres, une large revue de littérature dans laquelle nous adoptons une approche pluridisciplinaire. En effet, nous y mobilisons des travaux en économie expérimentale, en psychologie sociale et en gestion des organisations, liée à l'influence des émotions morales, dans le contexte des dilemmes sociaux. Cette démarche nous a conduit à mettre en perspective les résultats empiriques qui se sont dégagés dans ces différents champs scientifiques, de façon à offrir une vision d'ensemble du rôle des émotions morales, et en particulier la gratitude, qui permet d'envisager ces émotions comme un système d'outils diversifiés et complémentaires d'aide à la décision économique.

Dans le troisième chapitre, nous nous sommes ensuite interrogés sur la manière dont les économistes ont pris en compte certains des mécanismes psychologiques qui sont associés au rôle économique des émotions morales. A travers une revue non exhaustive des travaux issus de l'économie comportementale, nous montrons ainsi comment l'intégration progressive du rôle spécifique des processus et mécanismes affectifs, ayant abouti à la création des modèles dits de « préférences sociales », a permis de développer des modélisations proposant une représentation plus fidèle des comportements économiques observés expérimentalement, notamment via un traitement particulier de l'influence des émotions.

Le chapitre quatre est ensuite consacré à une présentation extensive du jeu du bien public, ainsi que des principaux résultats expérimentaux qui lui sont associés dans la littérature. Ce chapitre a ainsi pour finalité de justifier l'utilisation du cadre conceptuel que constitue le jeu du bien public, dans sa variété de déclinaisons, pour réaliser l'analyse formelle du rôle des émotions morales dans les dilemmes sociaux. En particulier, nous expliquons en quoi la conception même de ce jeu, notamment lorsqu'une phase de récompense lui est adjointe, est susceptible de déclencher l'émergence de l'émotion de gratitude chez les agents.

Enfin, ce travail de thèse aboutit, dans les cinquième et sixième chapitres, à l'élaboration d'un modèle original en économie comportementale, qui vise à analyser la façon dont l'émergence des émotions morales peut influencer, à la fois en statique et en dynamique, le choix stratégique des individus dans le jeu du bien public. En particulier, nous nous intéressons à la manière dont l'ajout d'une phase de récompense peut renforcer la coopération entre les partenaires et les aider à développer de bonnes relations interpersonnelles. Lors de ce travail de modélisation, nous adoptons des outils méthodologiques variés consistant, d'une part, en une approche analytique rigoureuse des résultats de notre modèle et, d'autre part, en une analyse recourant à la simulation ayant vocation à illustrer le fonctionnement des différents mécanismes psychologiques pris en compte dans notre conception.

Deuxième partie .

Le rôle des émotions morales dans la prise de décision économique

1. Les émotions morales : un système d'outils au service de la décision

1.1. Introduction

La thématique des émotions connaît depuis plusieurs années un regain d'intérêt au niveau scientifique à la fois sur le plan philosophique, neurobiologique, psychologique, sociologique, économique, politique ou même psychanalytique. Les émotions ont un intérêt particulier pour les économistes (Petit, 2009, 2015), pour les gestionnaires (van Hoorebeke, 2008) et au sein des organisations (Barsade et Gibson, 2007) : elles jouent un rôle prépondérant lors des processus de négociation ou de coopération entre partenaires individuels ou institutionnels (consommateurs, entreprises, salariés, gouvernements) ; elles modulent également la prise de décision individuelle ou collective, dans le cas notamment de situations de risque ou d'incertitude (achats d'actifs financiers, créations d'entreprise, management) ; elles sont enfin présentes au quotidien dans les activités de consommation (l'addiction), dans la relation au travail (stress, performance, créativité) ou au sein de l'organisation de l'entreprise (relations d'autorité ou culture d'entreprise).

Dans la littérature, certaines émotions comme celles du regret, de la déception, de l'envie ou de la colère ont été abondamment citées, en raison notamment du rôle primordial que ces émotions ont vis-à-vis de la prise de décision individuelle. Au sein d'une littérature plus récente, notamment dans le cadre de la négociation, certains travaux s'interrogent davantage au rôle des émotions sociales ou morales comme la colère, la culpabilité ou la honte (Morris et Keltner, 2000). La colère, par exemple, qui est suscitée par un événement déplaisant et perçu comme intentionnel, peut avoir des effets délétères sur la conduite d'un processus de négociation ou, au contraire, renforcer le pouvoir de négociation d'un des acteurs. Plus généralement, les émotions morales ont un impact crucial sur le respect des normes ou des valeurs définies par la société (honnêteté, altruisme, réciprocité, générosité, équité) ; elles modifient également les phénomènes de confiance entre les agents. De nombreux travaux en psychologie des émotions montrent en particulier que la prise de décision morale – conduisant à l'adoption de normes sociales ou morales – n'est pas le seul fruit d'un processus rationnel (Haidt, 2001 ; Valdesolo et DeSteno, 2006). D'après ce modèle, un processus intuitif émotionnel fonctionne en tandem avec des processus cognitifs conscients pour moduler la prise de décision morale. Les émotions morales auraient un intérêt tout particulier dans l'édification de comportements prosociaux (coopération, altruisme, confiance, etc.) qui sont bénéfiques à la société.

Dans ce chapitre, nous proposons une revue de la littérature en économie et en psychologie sur le rôle des émotions morales dans la prise de décision économique en présence d'interaction sociale. Nous mettons tout d'abord en perspective les spécificités des émo-

tions morales identifiées par la théorie des émotions (2). Nous analysons ensuite comment les émotions morales négatives – comme la culpabilité, la honte, l'embarras mais aussi la colère, l'indignation, le dégoût ou le mépris – et positives – comme la gratitude, la fierté, l'admiration ou l'empathie – conduisent à l'émergence et/ou au maintien des comportements prosociaux (3).

1.2. Les enseignements de la théorie des émotions morales

1.2.1. Qu'est-ce qu'une émotion ?

En psychologie, il n'y a pas de consensus sur la manière dont on peut définir ce qu'est une émotion. Il existe plutôt une pluralité de définitions, dont le recoupement permet de saisir les divers aspects qui les caractérisent. Les émotions sont des phénomènes internes qui peuvent parfois être observés au travers de certaines expressions ou de certains comportements (Niedenthal, Krauth-Gruber et Ric, 2006). Les émotions appartiennent à la catégorie des affects, un terme générique englobant notamment les émotions et les humeurs (Frijda, 1986). Une humeur est définie comme un état affectif diffus, de faible intensité, qui ne résulte pas nécessairement d'une cause facilement et spécifiquement identifiable. L'expérience d'une humeur peut persister sur des périodes pouvant aller de quelques minutes à quelques semaines. L'humeur est éprouvée de façon claire alors que sa survenance ou son extinction, ne sont pas forcément perçues de manière distincte. Comparativement aux humeurs, les émotions sont plus intenses, moins durables et ont généralement une cause bien définie. C'est lorsque l'on évalue une situation au regard de nos intérêts que nos émotions peuvent être déclenchées.

1.2.1.1. Une définition componentielle

La difficulté rencontrée par les psychologues pour définir clairement ce qu'est une émotion provient du fait que de multiples processus sont à l'œuvre lorsqu'une émotion survient. En effet, il peut s'agir de sentiments subjectifs, d'expressions physiques (faciales, vocales), d'une évaluation cognitive de la situation (à la source du ressenti émotionnel), d'une stimulation physiologique ou d'une tendance à entreprendre une action particulière (Frijda, 1986)¹. Ainsi, certains psychologues comme Keltner et Gross (1999), proposent une définition multi composantes de l'émotion. Pour ces auteurs, il est possible de définir les émotions comme des modes de perception biologique, des expériences, des états physiologiques, des actions et des communications qui sont épisodiques, de court-terme, et qui interviennent en réponse à des situations (challenges ou opportunités) qui nous concernent sur le plan physique ou social.

Ainsi, une émotion est ressentie à la suite d'un évènement déclencheur ou stimuli (une menace, une injustice, une maladresse...) et peut se manifester de plusieurs manières. L'émotion se traduit, tout d'abord, par une modification de l'état physiologique de l'individu, à travers l'activation de certaines zones du cerveau qui peuvent être différentes

1. Lorsque l'on ressent de la colère, on peut se sentir tendu, froncer les sourcils, plisser les yeux, interpréter la situation à l'origine de l'état émotionnel de colère comme injuste, négative, et contrôlable, ressentir une forte excitation, et préparer ses poings pour être prêt à attaquer la personne responsable de la situation.

selon l'émotion en question. Certaines émotions peuvent entraîner une modification de l'état physiologique qui sera source de plaisir (joie, gratitude, fierté, etc.), tandis que d'autres émotions sont à l'origine d'une modification de l'état physiologique synonyme de souffrance ou de désagrément (la tristesse, la colère, la culpabilité etc.). Il est donc possible de classer la perception des émotions que l'on ressent selon une échelle bi polaire que les psychologues appellent la valence et que l'on peut rapprocher du concept d'utilité employé en économie. On attribue une valence positive aux états émotionnels à caractère plaisant (joie), et une valence négative aux états émotionnels désagréables à ressentir (tristesse). L'impact physiologique d'un stimulus émotionnel peut également s'effectuer selon une autre dimension, le degré d'activation physiologique (arousal) ou degré d'implication émotionnelle. Celui-ci correspond à une mesure de l'intensité des effets psychophysiologiques des différentes émotions. Il peut être élevé (comme dans le cas de la colère) ou faible (comme dans le cas de l'ennui).

La survenance d'une émotion s'observe également via les comportements et les actions qu'elles incitent à entreprendre. On parle alors de « tendances à l'action » (Frijda, 1986). De nombreuses tendances à l'action ont été identifiées. Par exemple, une personne en colère peut avoir tendance à désirer se venger, l'émotion de culpabilité pousse un individu à réparer une erreur commise, la tendance à l'action de la honte est la dissimulation ou le repli sur soi. On parle ici de « tendances à l'action » et non pas d'actions spécifiques, chaque émotion pouvant donner lieu à une pluralité de comportements qui permettent une adaptation aux différentes situations dans lesquelles peut apparaître une même émotion : par exemple, la colère ne poussera pas à la même réaction si elle est dirigée à l'encontre d'un proche ou d'un inconnu. Ces tendances à l'action sont un aspect important des processus qui sont à l'œuvre lors d'un ressenti émotionnel car, selon Frijda (1986), l'émotion dispose d'une dominance vis-à-vis du contrôle du comportement, au sens où celle-ci passe avant les autres actions en cours. Cependant, même si les individus ne peuvent pas inhiber l'émotion elle-même, ils peuvent inhiber certaines réactions émotionnelles et gardent un certain contrôle de ces tendances à l'action. Niedenthal, Krauth-Gruber et Ric (2006) indiquent par exemple qu'une personne peut tout à fait ressentir de l'hilarité en se rappelant une blague lors d'une cérémonie de funérailles et pourtant se retenir de pouffer de rire, en affichant plutôt une expression de tristesse (même si parfois l'émotion est trop puissante, fou-rire).

Les tendances à l'action ne sont pas les seuls indices permettant de percevoir l'émergence d'un état émotionnel chez un individu. Il existe, en effet, de nombreux moyens d'expression qui sont généralement utilisés, consciemment ou inconsciemment, par les êtres humains. Il peut s'agir d'expressions posturales (croisement des bras ou tête baissée), vocales (intensité de la voix et vocabulaire), comportementales (tremblements ou diminution de la présence sociale) et même faciales (pleurs, grimaces, sourires). Ces expressions témoignent, aux yeux des autres, qu'un individu est dans tel ou tel état affectif. Cependant, comme dans le cas des tendances à l'action, les individus ont une capacité de contrôle sur ces moyens d'expression, comme un joueur de poker qui cherche à ne pas laisser transparaître ses émotions.

L'émotion intervient enfin sur le plan cognitif. En effet, la survenance d'une émotion entraîne une interruption des activités automatiques en cours, l'esprit se concentrant alors sur le contexte dans lequel l'émotion a émergé. Les ressources attentionnelles se focalisent

sur le traitement de l'information et fournissent une évaluation du contexte associé à l'émotion. Cette évaluation cognitive peut avoir un rôle au niveau du type d'émotion ou de l'intensité de la réaction émotionnelle résultant d'un évènement particulier. Par exemple, l'émotion de gratitude survient lorsque que l'on se rend compte, en partie à travers une évaluation réalisée sur le plan cognitif, que quelqu'un a fait quelque chose de bénéfique pour nous de manière intentionnelle et volontaire. De plus, l'aspect cognitif du processus émotionnel peut également entrer en jeu dans la mise en œuvre (ou dans la retenue) de la tendance à l'action (ou de l'expression de l'émotion) qui peut être évaluée comme souhaitable (ou, au contraire, indésirable) au regard du contexte particulier dans lequel l'émotion émerge. L'aspect cognitif des émotions a enfin une importance toute particulière vis-à-vis du fait qu'il permet une possibilité d'anticipation de la part des agents. A travers l'évaluation cognitive de la situation dans laquelle ils se trouvent, les agents peuvent ainsi susciter l'aide de leur mémoire et anticiper les émotions susceptibles de survenir. Comme le soulignait Adam Smith (1759), les émotions anticipées ou « imaginées », bien que non réellement ressenties, sont aussi capables de peser dans le processus de décision (moral) des agents. L'anticipation – via la recherche du ressenti de certaines émotions (positives) ou, au contraire, par tentative d'évitement (dans le cas des émotions négatives) – module en particulier la prise de décision.

1.2.1.2. Typologie

Dresser une typologie des émotions complète et pertinente est un exercice délicat en raison de la multitude des émotions répertoriées et de l'existence de frontières relativement floues entre les différentes émotions. Il est cependant possible en suivant Elster (1998) de proposer une typologie qui soit pertinente vis-à-vis de l'analyse économique. Dans sa typologie, Elster (1998) propose ainsi de regrouper les émotions selon cinq catégories. La première d'entre elles est celle des émotions dites sociales incluant les émotions de colère, de haine, de culpabilité, de honte, de fierté, de prétention, d'admiration et d'affection. Ces émotions peuvent être dirigées à l'encontre de nous-même ou à l'encontre d'une autre personne et concernent les comportements adoptés ainsi que le caractère (personnalité) même de la personne cible qui est à l'origine de l'émergence de l'émotion. Les émotions sociales impliquent également une évaluation positive ou négative de la personne cible. Ces émotions peuvent avoir un rôle prépondérant dans l'analyse économique lorsque l'on considère les situations d'interaction sociale telle que les situations de coopération, de négociation, de confiance, etc.

La typologie d'Elster (1998) intègre également la catégorie des émotions contrefactuelles, qui sont générées par la comparaison entre ce qui s'est passé et ce qui aurait pu arriver. Il s'agit, en particulier, des émotions de regret, de réjouissance (jubilation), de déception, d'exultation. Ces émotions sont en rapport étroit avec l'analyse économique, notamment dans un contexte d'incertitude et jouent un rôle important dans la modélisation du processus de décision des agents, en particulier en finance (Zeelenberg et Pieters, 2007).

Elster (1998) regroupe ensuite les émotions issues de ce qui pourrait se passer dans le futur, incluant la peur et l'espoir ; les émotions émergeant à la suite d'évènements, bons ou mauvais, qui ont déjà eu lieu telles que la joie et la tristesse (chagrin) ; ainsi que les émotions déclenchées par l'évaluation de ce que les autres possèdent : l'envie, l'indignation, la malveillance ou encore la jalousie. De plus, selon Elster, certaines émotions

n'appartiennent à aucune catégorie spécifique comme, par exemple, le mépris, le dégoût ou l'amour romantique. Enfin, il existe également dans cette typologie des émotions controversées ou qui sont à la frontière de plusieurs catégories comme la surprise, l'ennui, le désir sexuel, la jouissance, l'anxiété et la frustration.

La typologie d'Elster (1998) permet de regrouper l'étude des émotions de façon utile pour l'économiste. Elle n'est cependant pas exhaustive et ne permet pas, en particulier, de mettre en lumière le rôle majeur qu'ont les émotions dites « morales » dans le processus de décision économique. L'intérêt de ces émotions morales est à mettre en parallèle avec l'émergence d'une littérature abondante sur les comportements sociaux ou moraux (Serra, 2007).

1.2.2. La spécificité des émotions morales

Haidt (2003) définit les émotions morales comme les émotions qui sont « *intrinsèquement liées aux intérêts et au bien-être de la société dans son ensemble, ou au moins au bien-être d'un autre individu que l'agent qui les ressent* » (p. 276). Les émotions morales sont donc suscitées par des situations sociales impliquant la présence et l'intervention d'autrui. Elles présentent cependant des caractéristiques générales qui les différencient des émotions sociales (2.2.1). Une clé de lecture utile permettant d'identifier les spécificités des émotions morales et leur impact différencié sur l'adoption de comportements prosociaux consiste à séparer celles qui sont à valence négative (2.2.2) de celles qui procurent un bien-être ou une utilité positive à l'individu (2.2.3). Les mécanismes d'empathie ou de sympathie doivent cependant être traités à part (2.2.4). Nous traitons ensuite des spécificités liées à la mesure et à l'induction des émotions morales (2.2.5). Nous dressons enfin les enseignements principaux de la littérature sur les émotions morales (2.2.6).

1.2.2.1. Les caractéristiques générales

Les émotions morales peuvent apparaître lorsque les individus interagissent ensemble ou survenir lorsqu'une violation morale est perçue par les individus. L'intérêt pour ces émotions est de plus en plus grand chez les économistes car elles ont un rôle primordial dans la prise de décision des individus comme l'anticipait Adam Smith dans la *Théorie des Sentiments moraux* (1759). Certains psychologues formulent ainsi l'hypothèse que dans certaines interactions sociales, plus que notre capacité de raisonnement ou notre capacité cognitive, ce serait le ressenti d'émotions morales qui nous dicterait notre conduite (Vélez Garcia et Ostrosky-Solis, 2006 ; Kédia, 2009). Au sein de la famille des émotions morales, on retrouve la honte, la culpabilité, l'embarras, la colère, le mépris et le dégoût (des émotions à valence négative) mais aussi, la fierté, la gratitude et l'admiration (des émotions à valence positive). Pour de nombreux psychologues, l'empathie, enfin, représente une composante à part de cette catégorie d'émotion (Kédia, 2009).

Haidt (2003) propose une typologie des émotions morales comprenant les émotions auto conscientes, les émotions d'évaluation d'autrui et l'empathie. Les émotions auto-conscientes sont celles de la honte, de l'embarras, de la culpabilité et de la fierté. Elles ont en commun d'être déclenchées par les événements interpersonnels négatifs ou positifs dont la cause est attribuée à soi (Kédia, 2009). Il est, par exemple, possible de ressentir de la culpabilité ou de la honte, après s'être comporté de façon opportuniste en profitant du fruit d'un effort collectif sans avoir contribué à sa réalisation, ou au contraire de

1. Les émotions morales : un système d'outils au service de la décision

ressentir de la fierté, du fait d'un profond investissement dans la réalisation de ce projet collectif. Ces émotions sont évoquées par l'auto-évaluation de soi et par la réflexion sur soi (Tangney, Stuewig et Mashek, 2007). Elles constituent des indicateurs précieux pour évaluer la façon dont nous sommes perçus par notre environnement social et nous aident ainsi à maintenir de bonnes relations avec les autres. En particulier, elles nous informent sur l'acceptabilité sociale ou morale de nos conduites, en agissant comme des punitions ou des récompenses affectives. Ces émotions d'auto-évaluation fonctionnent donc, d'après Tangney *et al.* (2007), comme un « baromètre émotionnel moral ». En particulier, lorsque l'on commet une transgression ou que l'on fait une erreur préjudiciable pour autrui, les émotions de culpabilité, de honte ou d'embarras sont susceptibles d'apparaître et induisent un désagrément, une désutilité. A l'inverse, lorsque notre comportement est socialement ou moralement adapté, les sentiments positifs de fierté ou d'approbation de soi-même en résultent, venant récompenser un comportement jugé opportun.

Les émotions morales auto-conscientes sont fondamentales pour les êtres humains dotés biologiquement d'un fort besoin d'appartenir à un groupe. Elles semblent ainsi conçues pour aider les individus à s'adapter à la vie de groupe en évitant les comportements susceptibles d'entraîner le mépris, la colère ou le dégoût chez les autres personnes (Haidt, 2003) ou alors en favorisant les comportements socialement et moralement adaptés capables de susciter l'admiration ou la gratitude auprès des autres. Ce sont, par ailleurs, ces autres émotions qui constituent la deuxième grande catégorie d'émotions morales, les émotions d'évaluation d'autrui. Lorsque l'on considère ces émotions, l'objet de l'évaluation n'est plus soi-même, celles-ci se focalisent plutôt sur les comportements adoptés par les individus qui nous entourent. Comme pour les émotions auto-conscientes, il existe une gamme d'émotions d'évaluation d'autrui à valence négative que Haidt (2003) nomme les émotions de condamnation d'autrui, intégrant les émotions de colère/indignation, de mépris et de dégoût, ainsi qu'une gamme d'émotions d'évaluation d'autrui à valence positive, comme les émotions de louange d'autrui, parmi lesquelles on retrouve notamment la gratitude et l'admiration.

D'après Kédia (2009), les émotions de condamnation d'autrui, la colère/indignation, le dégoût et le mépris sont des émotions morales, au même titre que les émotions qui incitent à adopter des comportements prosociaux, car elles constituent des mécanismes essentiels à la régulation des relations interpersonnelles et au maintien de l'ordre social. Ces émotions conduisent, en effet, à la vengeance, à l'ostracisme ou à l'humiliation vis-à-vis de l'objet de l'évaluation, c'est-à-dire la ou les personnes ayant commis une transgression. Ces émotions incitent ainsi à la mise en place de mesures coercitives de la part des individus qui les ressentent envers les individus qui en sont la source, du fait de leurs comportements socialement ou moralement inadaptés. Ceci conduit, notamment via *l'anticipation* de ces mesures, à décourager ces « mauvais » comportements. A l'inverse, les émotions de louange d'autrui surviennent lorsqu'on constate que les individus qui nous entourent se comportent de manière admirable ou réalisent une action qui vise à augmenter notre bien-être ou celui de la société dans son ensemble. Au même titre que les émotions de condamnation d'autrui impliquent une *réciprocité négative* de la part de ceux qui les ressentent envers ceux qui en sont la source, les émotions de louange d'autrui peuvent être à l'origine d'une *réciprocité positive* qui récompense l'adoption de comportements prosociaux. Ces émotions positives encouragent également les individus à créer des liens sociaux, à devenir de meilleures personnes elles-mêmes (Haidt, 2003) et à manifester la

volonté d'aider les autres. Les émotions de louange d'autrui sont susceptibles de renforcer les comportements prosociaux de ceux qui les ont réalisés, mais aussi de promouvoir ces comportements auprès de ceux qui les ressentent.

Au sein de ces deux sous catégories d'émotions morales, on peut faire un regroupement par valence pour présenter les différentes émotions et leurs caractéristiques, puisque les émotions morales à valence négative peuvent être vecteurs de mécanismes de punitions (orientés vers soi-même ou vers les autres) vis-à-vis de comportements moralement ou socialement inadaptés, tandis que les émotions à valence positive sous-tendent à l'inverse des mécanismes de récompenses (qui peuvent à nouveau être orientés vers soi-même ou vers les autres) en lien avec des comportements et des actions prosociales.

1.2.2.2. Les émotions morales à valence négative

1.2.2.2.1. La culpabilité, la honte et l'embarras

Dans la littérature en psychologie, les frontières entre les émotions auto conscientes à valence négative sont relativement floues. D'après Haidt (2003), dans certaines cultures non occidentales comme les cultures asiatiques, il n'existe pas de différences lexicales entre honte et embarras. Dans d'autres cultures non occidentales, il est même suggéré que l'émotion de culpabilité n'existerait pas et serait (plus ou moins) intégrée dans la honte ou l'embarras. Pour autant, il existe des nuances entre ces différentes émotions et une manière de pouvoir observer ces différences est de les présenter en comparant ces émotions les unes aux autres.

Honte versus culpabilité

La *honte* et la *culpabilité* sont des sentiments qui surviennent lorsqu'une personne réalise une action qui *viole une norme sociale ou morale*. La culpabilité serait cependant plus pertinente dans le cadre de violations de normes morales qui causent du tort à autrui². Tangney, Marshall, Rosenberg, Barlow et Wagner (1994) indiquent, par ailleurs, qu'en décrivant des situations où ils ressentent de la honte, les sujets de leur étude expriment plus d'inquiétude concernant l'évaluation des autres à propos d'eux-mêmes (comparativement à l'expérience de la culpabilité), tandis qu'en décrivant des situations où ils ressentent de la culpabilité, les sujets sont plus préoccupés sur l'effet qu'ils peuvent (à travers leur comportement) avoir sur les autres. Ainsi, l'égoïsme qui intervient lorsqu'on ressent de la honte amène à penser que celle-ci pourrait être plus prégnante que l'émotion de culpabilité lors d'une transgression d'une norme morale qui a lieu publiquement, puisque l'on est directement soumis au regard et au jugement des autres sur soi-même³.

La distinction la plus tranchée entre l'émotion de honte et l'émotion de culpabilité se situe au niveau de la différence de focus qui intervient lors du ressenti de ces émotions. Dans le cas de la honte, le focus est centré sur soi-même (« J'AI fait cette chose terrible » / « quelle horrible personne JE SUIS »), alors que dans le cas de la culpabilité, l'accent est plutôt mis sur le comportement adopté (« j'ai fait cette CHOSE terrible » / « quelle horrible CHOSE j'ai faite »). D'après Haidt (2003), l'émotion de honte ne serait donc

2. Cependant, d'après Tangney *et al.* (2007) la plupart des violations de normes morales telles que le fait de voler, de tromper, de refuser d'aider quelqu'un, etc., peuvent être à la fois citées comme entraînant un sentiment de honte par certaines personnes et comme entraînant de la culpabilité par d'autres.

3. Pour autant, d'après Tangney *et al.* (2007), les données empiriques peinent à montrer une telle différence.

1. Les émotions morales : un système d'outils au service de la décision

pas un régulateur de comment quelqu'un doit agir mais plutôt de comment quelqu'un doit être. En conséquence, dans les sociétés modernes occidentales, la honte est ainsi une émotion plus douloureuse, mettant en jeu la totalité de la personnalité de l'individu et altérant son estime de soi. Sur l'échelle de valence, la honte possède ainsi généralement une plus forte *intensité* négative que la culpabilité, la *désutilité* entraînée par le fait de ressentir de la honte pouvant donc être supérieure à celle provenant du fait de ressentir de la culpabilité.

Les émotions de culpabilité et de honte ont enfin des *tendances à l'action* bien différentes. La culpabilité, considérée comme l'émotion morale par excellence par de nombreux spécialistes des émotions, entraîne le plus souvent une aide des personnes en détresse (permettant de ne pas ressentir de culpabilité vis-à-vis de ces personnes), une *réparation* du tort commis, ou encore des excuses ou des confessions. Cette tendance à l'action vise à atténuer ou à éliminer le sentiment de culpabilité et vise également au bon rétablissement de la relation mise en péril par le tort commis (Kédia, 2009 ; Haidt, 2003). A contrario, la honte, entraîne un désir de réduire sa présence sociale en se cachant ou en fuyant. La honte peut également être associée à un sentiment de colère et à de fortes réactions de vengeance, liées à un désir de défense de l'estime ou de l'image de soi (Tangney *et al.* 2007). Ainsi, là où le sentiment de culpabilité produit une attitude constructive de « réparation », le sentiment violent de honte a plutôt tendance à induire des comportements nuisibles au prolongement de la relation sociale. Par anticipation, la honte peut cependant inciter les individus à éviter les situations dans lesquelles ils pourraient ressentir cette émotion. Ainsi, la honte, ne possédant pas une tendance à l'action prosociale immédiate, peut s'avérer être un outil dissuasif à plus long terme.

Honte versus Embarras

L'*embarras* est une émotion très proche de celle de la honte. Cependant, Les émotions auto-conscientes sont suscitées par une auto-évaluation de soi qui est effectuée vis-à-vis des normes morales et des conventions sociales. Or, d'après Haidt (2003), les sociétés occidentales font partiellement la différence entre les problèmes d'ordre moraux (quand quelqu'un fait du mal à une autre personne, les questions concernant le droit, la justice) et les problèmes en rapport avec les conventions sociales (le choix des habits, de la nourriture, l'hygiène). Ainsi, l'auteur explique que l'embarras serait plus ressenti dans le cadre d'une violation d'une convention sociale, tandis que la honte serait plus en rapport avec des violations d'ordre moral. Il précise également que l'embarras, qui est ressenti dans des situations qui mettent moins en danger la présentation de soi auprès des autres, est moins violent que le sentiment de honte. Enfin, il relie le sentiment d'embarras avec le fait d'évoluer à proximité d'un individu de plus haut statut que nous.

Pour Tangney *et al.* (2007), l'embarras est une émotion un peu moins en rapport avec la moralité que ne peuvent l'être la honte et la culpabilité. D'après Miller (1992, 1995 cité dans Tangney *et al.*, 2007) l'embarras est défini comme un état aversif de mortification, de confusion et de chagrin suivant une situation sociale difficile (embarrassante) ayant lieu publiquement telle que les situations dans lesquelles une personne se comporte de manière maladroite, distraite ou malchanceuse. L'embarras est par conséquent en lien avec des situations qui mettent moins en danger la présentation de soi auprès des autres et provoque, en conséquence, une désutilité moins forte que celle de la honte. Malgré son rapport plus privilégié avec les conventions sociales qu'avec les normes morales (comparativement au

sentiment de honte), Tangney *et al.* (2007) indiquent que le sentiment d'embarras pourrait avoir des implications pour les comportements moraux. Certaines recherches ont ainsi montré que les personnes embarrassées ont tendance à agir de manière conciliante pour gagner l'approbation des autres et la réintégration au sein d'un groupe. Le sentiment d'embarras peut donc pousser les individus à se comporter de manière socialement acceptable, ce qui peut inclure le respect des normes morales qui s'imposent à eux. De plus, comme dans le cas de la honte ou de la culpabilité, l'anticipation des situations embarrassantes pousse les individus à respecter les conventions sociales ou morales et guider *ex ante* leur processus de décision.

Ainsi, en résumé, les sentiments de honte, de culpabilité et d'embarras sont des émotions qui surviennent lorsqu'on commet une transgression d'une norme morale ou sociale, chez les personnes qui réalisent cette transgression. Elles sont à valence négative et impliquent en conséquence une désutilité. Ces émotions peuvent donc peser, à travers l'aspect cognitif du processus émotionnel, sur le processus de décision des individus, lorsque ceux-ci anticipent les situations où ils risquent de les ressentir. De plus, la culpabilité et l'embarras possèdent des tendances à l'action qui vise à un renfort d'une relation sociale mise en péril et à une réintégration dans le groupe par l'intermédiaire de comportements ou d'actions prosociales.

1.2.2.2. La colère, le dégoût et le mépris

Les émotions de condamnation d'autrui proviennent d'une évaluation négative du comportement ou des actions entreprises par les individus qui nous entourent. Elles jouent un rôle primordial dans le maintien du respect des normes morales et sociales en incitant à la mise en place de mécanismes punitifs (de vengeance, d'ostracisme et de moquerie) à l'encontre de ceux qui violent ces normes. On retrouve à l'intérieur de ce groupe d'émotions, la colère et l'indignation, ainsi que le dégoût et le mépris.

La colère et l'indignation

L'émotion de *colère* est une émotion primaire, présente chez tous les individus, dans toutes les cultures du monde. Elle est commune à la fois aux êtres humains et aux autres espèces vivantes telles que les rats ou les chiens, chez qui elle survient en cas de frustration. Chez les êtres humains, la colère survient également dans d'autres situations, comme lorsque l'on est insulté, frustré, incommodé, ou blessé (physiquement ou psychologiquement). La colère surgit notamment lorsque l'on est victime d'un préjudice, surtout lorsque celui-ci est perçu comme injuste (Ortony, Clore et Collins, 2003), qu'il soit effectif ou potentiel, et que l'on en attribue la responsabilité et l'intention à un offenseur identifié. On peut retrouver l'influence de la colère dans de nombreuses situations économiques. Entre autres, la colère peut être ressentie lors d'un processus d'échange entre deux partenaires, lorsque l'un d'eux fait une proposition jugée injuste par l'autre. Consécutivement à l'émergence de cette émotion, les principales tendances à l'action qui ont été identifiées sont la recherche de vengeance et la demande de compensations ou de réparations (Kédia, 2009). Par exemple, dans une situation de négociation, l'un des partenaires peut décider de se venger en rejetant une offre qu'il estime injuste. L'émotion de colère est donc source de réciprocité négative vis-à-vis de la source de l'émergence de cette émotion. D'après Haidt (2003), cette tendance à l'action peut sembler à la fois égoïste et antisociale. La colère possède cependant d'autres aspects qui lui confèrent les caractéristiques d'une émotion morale.

1. Les émotions morales : un système d'outils au service de la décision

La relation entre la colère et la morale est en fait sans doute réciproque (Watts, 2007). La colère reflète la sensibilité morale d'un individu, au sens où une personne amoral est sans doute incapable d'une réaction émotionnelle qu'on peut nommer de la « colère ». En effet, la colère survient également en réponse à des événements ou des comportements représentant une violation des normes morales, auquel cas, le préjudice n'a pas besoin d'être personnellement subi. Cette émotion contribue donc à forger notre sensibilité morale et mobilise des comportements prosociaux. La colère « vertueuse » peut en particulier émerger à la suite de l'observation d'un mauvais comportement, dirigé à l'encontre d'une tierce personne (Tangney *et al.*, 2007). Cette forme de colère, qui est également désignée sous le terme d'*indignation*, recouvre une fonction morale puisqu'elle motive et active la réparation d'injustice. L'individu responsable de la transgression, objet de l'indignation, subit en effet en retour un préjudice matériel (rupture d'un contrat, sanction monétaire) ou psychologique. Dans ce dernier cas, le transgresseur, confronté à la colère d'un autre individu, peut ressentir une douleur psychophysiologique sous la forme de méta émotions telles que la peur ou la culpabilité induite par la colère d'un tiers. Ces punitions peuvent, par anticipation, modifier le processus de décision des individus en faveur de comportements évitant de susciter la colère d'autrui.

Le dégoût

Le *dégoût* se réfère, selon Darwin (1872), à quelque chose de révoltant, principalement en rapport avec le sens du goût (réellement expérimenté ou seulement imaginé) mais qui peut également, en second lieu, être en rapport avec toute chose qui cause une sensation similaire à travers l'utilisation de l'odorat, du toucher ou de la vue. Cette définition peut être élargie, d'une part aux idées ou aux actes qui impliquent une utilisation dégoûtante de son corps (incestes, abus d'enfants, pratiques sexuelles déviantes, abus de drogue, etc.) et, d'autre part, aux violations purement sociales et morales (Rozin, Haidt et McCauley, 2000). L'hypocrisie, la trahison ou la cruauté sont, par exemple, des comportements qui font émerger l'émotion de dégoût, plutôt que l'émotion de colère ou de haine (Miller, 1998).

Au contraire de la colère, qui implique une tendance à vouloir attaquer ou se venger de celui qui en est à la source, la tendance à l'action de l'émotion de dégoût est plutôt de vouloir se protéger. Elle consiste, en effet, à se retirer et à éviter la source du dégoût, que celle-ci soit, par exemple, un plat de nourriture avarié ou une personne ayant trahi notre confiance. D'après Haidt (2003), cette tendance serait particulièrement utile aux individus lorsque l'on considère l'aspect sensoriel de l'émotion de dégoût, car elle permet d'éviter la consommation de nourritures potentiellement contaminées par des bactéries pouvant être mortelles. L'aspect moral de cette émotion peut quant à lui être à l'origine de tendances à l'action recouvrant un rôle prosocial. En rejetant en particulier les individus qui déclenchent le dégoût socio moral (ostracisme), la société met en place une structure de punitions et de récompenses qui agit comme une puissante force de dissuasion vis-à-vis des comportements culturellement (socialement et moralement) inappropriés. Ainsi, même si la tendance à l'action du dégoût est moins agressive que celle de la colère, elle n'en constitue pas moins un mécanisme de réciprocité négative, au travers de punitions telles que l'ostracisme, pour ceux qui en sont à la source. Ces individus, afin d'éviter de subir une mise à l'écart de la part des autres individus ou de ressentir des émotions telles que la honte ou l'embarras, peuvent, par anticipation, modifier leur choix de comportement.

Le mépris

Comme la colère et le dégoût, l'émotion de *mépris* est le produit d'une évaluation négative des autres. Dans la littérature en psychologie, le mépris pour quelqu'un implique de le regarder de haut et de se sentir moralement supérieur à celui-ci. Le mépris serait en particulier en rapport avec la position hiérarchique relative qui existe entre les individus (Tangney *et al.*, 2007).

Cette émotion est sans doute un peu moins en rapport avec la notion de morale que peuvent l'être la colère ou le dégoût. Néanmoins, la tendance à l'action du mépris est de moquer et de traiter avec moins de respect les personnes qui en sont la source, c'est-à-dire ceux que l'on considère comme moralement inférieurs, que ce soit du fait de leur position hiérarchique comme du fait d'un comportement moralement méprisable (tel que le fait de ne pas être à la hauteur du statut ou du niveau de prestige que l'on revendique). Cette tendance à l'action s'interprète ici aussi comme une punition proactive exercée à l'encontre de la personne méprisée. Par exemple, au sein de l'entreprise, certains individus effectueront leur niveau d'effort maximum afin de limiter les risques d'encourir la réprobation et le mépris de leurs collègues ou de leur employeur.

Les émotions de condamnation d'autrui ont donc un rôle primordial dans le maintien du respect des normes morales et sociales puisqu'elles incitent à la mise en place de mécanismes punitifs (de vengeance, d'ostracisme et de moquerie) à l'encontre de ceux qui violent ces normes. L'anticipation, de la part des transgresseurs potentiels, de l'émergence des sentiments de colère/indignation, de dégoût et de mépris et les tendances à l'action qui en résultent, peut être capable de dissuader la transgression et, ainsi, favoriser l'adoption de comportements prosociaux.

1.2.2.3. Les émotions morales à valence positive

Les émotions à valence négative ont été très abondamment étudiées dans la littérature en psychologie. A contrario, les émotions positives, comme la joie, l'excitation, l'amour ou l'admiration, sont encore relativement peu connues. Dans le cas des émotions morales à valence positive, on retrouve la distinction entre les émotions auto-conscientes, qui concernent essentiellement la fierté, et les émotions d'évaluation d'autrui, comme la gratitude et l'admiration.

1.2.2.3.1. La fierté

La fierté est déclenchée par les événements interpersonnels positifs que l'on peut attribuer à soi. La fierté est ainsi « l'opposée » de l'émotion de honte : cette dernière est éprouvée lorsque l'on pense avoir échoué à satisfaire une représentation idéale de soi tandis que la fierté est, au contraire, ressentie lorsqu'on pense avoir été à la hauteur de cette représentation idéale.

Mascolo et Fisher (1995) définissent la fierté (morale) comme une émotion générée lorsque l'on se rend compte que l'on est responsable de l'adoption d'un comportement socialement valorisé, ou que l'on est soi-même quelqu'un de valorisé socialement par les autres. La fierté sert à encourager les comportements futurs qui se conforment aux standards sociaux de valeur ou de mérite. Cet encouragement s'effectue par anticipation de l'émergence de l'émotion de fierté, qui est une émotion agréable à ressentir. Le sentiment

1. Les émotions morales : un système d'outils au service de la décision

de fierté, du fait qu'il incite à se conformer, voire à faire mieux que ce qu'imposent les normes moralement pertinentes (ou du fait qu'il entrave les impulsions à agir de manière immorale), recouvre une fonction relativement importante en récompensant et en renforçant l'engagement des individus aux normes morales et sociales (Tangney *et al.*, 2007). Il y a par ailleurs un lien fort entre la honte et la fierté vis-à-vis de ces normes, la conformité aux normes sociales et morales étant provoquée par des récompenses et des punitions informelles (Scheff, 1988). Ainsi, parallèlement au cas de la culpabilité, de la honte ou de l'embarras, mais cette fois sous un versant positif, la fierté serait une sorte de récompense intériorisée liée à l'adoption et à la réalisation de comportements moralement pertinents. Par conséquent, ce mécanisme de récompense peut s'avérer incitatif vis-à-vis de l'adoption de ces comportements ⁴.

1.2.2.3.2. La gratitude et l'admiration

La gratitude

Dès la fin du 18^{ème} siècle, Adam Smith insistait sur le rôle primordial de la gratitude dans les relations sociales. Dans la *Théorie des Sentiments Moraux* (1759), la gratitude est une émotion sociale fondamentale permettant de fournir aux agents des orientations vis-à-vis de leurs jugements et de leurs comportements moraux. La gratitude est ainsi primordiale pour le fonctionnement de la vie sociale, du fait qu'elle permet de promouvoir une plus grande stabilité des relations au sein de la société. Malgré cet intérêt précoce d'Adam Smith, peu d'économistes ou même de psychologues se sont penchés, jusqu'au milieu du 20^{ème} siècle, sur l'étude de cette émotion (McCullough, Kilpatrick, Emmons et Larson, 2001).

La gratitude est définie comme l'émotion à valence positive que l'on ressent « *lorsqu'une autre personne a effectué ou a tenté d'effectuer un acte, de manière intentionnelle et volontaire, entraînant un gain particulièrement précieux (un don d'argent, de l'aide), qui représente un coût pour la personne qui le réalise, sans compensation financière (immédiate) de la part de celui qui le reçoit* » (McCullough *et al.*, 2001) ⁵. En tant qu'émotion à valence positive très agréable à ressentir, la gratitude contribue au bien-être personnel, à l'engagement civique et à la satisfaction spirituelle (Emmons et Crumpler, 2000). On lui attribue également quatre fonctions prosociales : la gratitude est (1) un baromètre moral, (2) une incitation à la réciprocité positive directe ou (3) indirecte et, enfin, (4) un mécanisme de « renforcement moral ».

4. Il convient cependant de distinguer l'émotion de fierté de la vanité (vantardise ou arrogance). La vanité se définit comme le fait de s'attribuer un accomplissement qui n'est pas le nôtre, qui n'est pas admirable, qui ne nous appartient pas complètement (Lea et Webley, 1997), ou encore le fait de s'accorder une valeur exagérée. Cet aspect de l'émotion de fierté, ou, au moins, en lien avec celle-ci, pose des soucis. En effet, l'arrogance et la vanité sont, d'après Lewis (1992), largement inadaptés, car les individus ressentant ces émotions ont tendance à déformer la réalité ou à l'inventer pour se mettre en avant, ce qui peut impliquer des problèmes relationnels.

5. Il convient de la différencier du sentiment de dette. Ce sentiment est, contrairement à la gratitude, désagréable à ressentir, il place le focus sur l'obligation de rembourser celui qui a agi intentionnellement dans notre intérêt et provoque la peur de ne pas être capable de rembourser ainsi que l'anxiété résultant de la recherche d'un moyen de pouvoir effectuer ce remboursement (McCullough *et al.*, 2001). Le sentiment de dette provient de la communication des exigences de retour (ou de remboursement) de la part de celui qui a fait quelque chose pour nous, et il résulte en une diminution du désir de vouloir l'aider (Watkins, Scheer, Ovnicek et Kolts, 2006).

1. La gratitude est une réaction émotionnelle liée à la perception que l'on obtient un bénéfice provenant d'une action de la part d'un autre agent ; elle fournit en ce sens une mesure de l'évaluation morale (un baromètre) de son propre comportement.
2. La deuxième fonction prosociale est induite par la tendance à l'action de la gratitude qui implique une forme de *réciprocité positive* (en « rendant la pareille » à celui qui est l'objet de notre gratitude). Ainsi, lorsque l'on ressent de la gratitude envers une personne qui a réalisé une bonne action à notre égard, cela nous incite à récompenser cette personne, en réalisant à notre tour une action prosociale en sa faveur. La gratitude servirait donc de « *moteur moral* », incitant les individus à faire preuve de *réciprocité positive*. D'après Trivers (1971), la gratitude fait ainsi intégralement partie du mécanisme émotionnel de *réciprocité altruiste*, encourageant les individus à rembourser leurs bienfaiteurs. En ce sens, la gratitude, favorisant un mécanisme de récompense, serait l'exact opposé de la colère, qui a tendance à sanctionner une attitude sociale ou morale inadaptée. Cette tendance à l'action revêt un rôle particulièrement important dans les relations sociales, la gratitude (au même titre que la culpabilité et l'amour) ayant un rôle pivot dans la construction d'une relation de confiance entre les individus, en encourageant à adopter des comportements qui soutiennent (et cimentent) celle-ci, même lorsque ces actions représentent un coût à court terme pour ceux qui les entreprennent (Barlett et DeSteno, 2006 ; Frank, 1988). La gratitude peut donc participer au maintien et au renforcement des bonnes relations existantes entre les individus, et donc avoir un rôle dans la constitution des réseaux et dans les situations de négociation.
3. La gratitude peut également inciter ceux qui la ressentent à entreprendre des actions prosociales ne se limitant pas à des comportements orientés vers le bienfaiteur (McCullough *et al.*, 2001). Elle induit également une *réciprocité indirecte* lorsqu'un individu, ressentant de la gratitude suite à un comportement prosocial réalisé par un bienfaiteur, décide à son tour de se comporter de manière prosociale à l'encontre d'une tierce personne.
4. En encourageant et en motivant les bienfaiteurs, qui sont à la source de l'émergence de la gratitude, à persister dans leurs comportements prosociaux, la gratitude exerce une fonction de « *renforcement moral* ». C'est le cas lorsque le mécanisme de récompense issu de l'émergence de l'émotion de gratitude, permet (éventuellement à travers la naissance d'un sentiment de fierté) d'inciter les bienfaiteurs à continuer à adopter ces comportements ⁶.

L'Admiration

A côté de la gratitude, il existe une autre émotion au sein de la famille des émotions de louange d'autrui. Il s'agit de l'admiration, terme le plus proche correspondant à la traduction du mot anglais « *elevation* » chez les auteurs anglo-saxons. D'après Haidt (2003), cette émotion survient lorsque l'on est témoin d'une manifestation de la beauté morale, c'est-à-dire d'un acte profondément moral tel que la charité, la gentillesse, la

6. On retrouve donc un mécanisme similaire, mais opposé, à celui qui est en place avec les émotions morales négatives de condamnation d'autrui, la colère ayant ainsi un impact prosocial sur la personne qui est à l'origine de l'émergence de l'émotion chez autrui.

1. Les émotions morales : un système d'outils au service de la décision

loyauté ou le sacrifice. Haidt (2003) indique que ce sentiment semble être l'opposé du sentiment de dégoût qui est lui déclenché lorsque l'on est témoin d'un acte de dépravation morale. La tendance à l'action associée à cette émotion est le ressenti d'un affect positif (évaluation positive) envers celui qui en est à l'origine. Cependant, à la différence de la gratitude, dont la tendance à l'action se concentre particulièrement sur le bienfaiteur qui a agi en notre faveur, la tendance à l'action de l'admiration serait plus générale et consisterait à vouloir devenir soi-même une meilleure personne. Cette émotion peut ainsi conduire à récompenser l'individu admiré mais est surtout à l'origine d'un désir d'adopter soi-même ce type de comportement moral. Ceci indiquerait une capacité pour la gratitude, et à plus forte raison pour l'admiration, de créer des mécanismes de réactions en chaînes (ou de « spillovers » de comportements moraux) qui peuvent être bénéfiques pour l'ensemble de la société.

1.2.2.4. Empathie /sympathie

Dans la typologie des émotions donnée par Haidt (2003), on distingue une catégorie un peu particulière, celle des émotions de souffrances d'autrui, qui regroupe deux émotions relativement proches : les émotions d'empathie et de sympathie. Ces deux émotions impliquent une prise en compte de la situation, des émotions, ainsi que de la souffrance d'autrui. Historiquement, l'émotion de souffrance d'autrui qui s'est, la première, dégagée dans la littérature est la sympathie, qualifiée de fondation de la moralité par Adam Smith (1759), David Hume (1896) et Jean Piaget (1973). Haidt (2003) souligne que ces auteurs avaient caractérisé la capacité des êtres humains à se sentir mal lorsqu'ils voient un autre être humain souffrir, et que parfois, ces sentiments poussent les individus à venir en aide aux personnes qui sont ostensiblement en détresse. L'émotion d'empathie intègre la littérature un peu plus tardivement, le terme ayant été inventé en 1909 par Titchener comme une traduction du mot allemand « *einführung* ». L'empathie correspond au processus conduisant à voir les événements de l'intérieur (Batson et Shaw, 1991). L'empathie peut être définie comme « *une réponse émotionnelle provenant de la condition ou de l'état émotionnel d'une autre personne et qui est congruente avec la condition ou l'état émotionnel de cette autre personne* » (Haidt, 2003). L'empathie est donc la capacité de pouvoir ressentir une ou des émotions, sur la base de la perception d'une ou de certaines émotions similaires ressenties par une autre personne ⁷.

D'avantage que des émotions purement discrètes, l'empathie et la sympathie sont des processus émotionnels à travers lesquels les individus sont amenés à ressentir des émotions telles que la colère, la tristesse ou la joie en fonction des émotions qu'ils perçoivent chez une autre personne. Tangney (2007) rapporte que, selon Feshbach (1975), la capacité à faire preuve d'empathie nécessite trois compétences : (a) l'aptitude cognitive à se mettre à la place de quelqu'un d'autre, (b) l'aptitude cognitive de reconnaître et de différencier les émotions ressenties par une autre personne et enfin (c) l'aptitude affective de ressentir personnellement une gamme d'émotions.

7. Les deux émotions de sympathie et d'empathie sont relativement proches, mais certains auteurs effectuent une distinction entre elles. D'après Tangney (2007), en effet, des chercheurs comme Eisenberg (1986) expliquent que l'émotion de sympathie implique un sentiment d'attention vis-à-vis de l'état émotionnel des autres, mais qu'au contraire de l'empathie, il n'implique pas le fait de ressentir viscéralement les mêmes émotions.

L'empathie ou la sympathie sont déclenchées par la perception de la souffrance ou de la tristesse chez une autre personne, que cette personne soit un proche ou qu'elle soit totalement étrangère. Elles conduisent généralement les personnes qui ressentent de l'empathie à aider, à reconforter et à tenter de diminuer la souffrance de l'autre. Ces processus émotionnels ont donc un rôle majeur vis-à-vis des comportements sociaux ou moraux puisque la capacité des individus à faire preuve d'empathie va conditionner leur capacité à adopter des comportements sociaux tels que d'aider une personne en détresse. En effet, d'après Kédia (2009), les individus qui possèdent une personnalité empathique, ceux qui sont donc plus susceptibles de réagir par la compassion et d'être touchés par la souffrance d'autrui, semblent avoir particulièrement tendance à adopter des conduites prosociales (donner de l'argent à des œuvres de charité, accepter de consacrer du temps à aider des personnes en difficulté)⁸.

Un autre aspect primordial du processus émotionnel empathique est qu'il permet, d'une certaine manière, de faire le lien entre les différents mécanismes de punitions et de récompenses développés par les autres familles d'émotions morales, et de contribuer à leur efficacité. En effet, d'après Tangney (2007), l'empathie est très fortement corrélée au sentiment de culpabilité et cette capacité à ressentir de l'empathie participerait aux efforts et aux comportements réparateurs de la part de la personne qui se sent coupable. De plus, d'après Kédia (2009), les psychopathes ou certains patients ayant subi des accidents causant des dommages dans certaines zones du cerveau, développeraient des comportements immoraux et asociaux tels que des comportements agressifs, du fait qu'ils ne seraient pas capables d'anticiper les émotions de condamnation d'autrui (colère, dégoût, mépris) qui en résulteraient. Ainsi, leur déficit de capacités empathiques, ne leur permettrait pas, d'une part, de se rendre compte du mal qu'ils vont faire, ce qui implique une absence de sentiment de culpabilité ou de honte, et d'autre part d'anticiper l'émergence de sentiments négatifs (colère, dégoût) chez les autres, à l'origine de mesures punitives. On voit ici le rôle qui lie l'empathie avec les émotions morales à valence négative (qu'elles soient auto-conscientes ou non). Cependant, l'empathie peut également intervenir au sein des familles d'émotions morales à valence positive. D'après Lazarus et Lazarus (1994), cités dans McCullough *et al.* (2001), la gratitude est une émotion empathique à part entière. Celle-ci nécessite en effet de pouvoir déterminer et anticiper l'intentionnalité du geste réalisé par le bienfaiteur. Cette capacité pourrait transiter par le processus émotionnel d'empathie. De plus, le bienfaiteur peut également être motivé dans son acte par l'anticipation, éventuellement à travers un processus empathique, de l'émergence de l'émotion de gratitude qui en résultera, et par l'anticipation d'un sentiment de fierté ou de bien-être (« warm-glow ») qu'il ressentira du fait d'accomplir un acte prosocial.

L'intérêt des psychologues et des économistes pour les émotions et leur impact sur le comportement des individus a nécessité la mise en place de nombreuses techniques afin d'obtenir des données empiriques. Le plus souvent, les chercheurs tentent d'observer et d'analyser l'influence d'une ou plusieurs émotions particulières, dans un contexte le plus neutre possible. Un des enjeux des procédures expérimentales, mises en place par les chercheurs, est donc de pouvoir répondre aux deux questions suivantes : comment

8. L'empathie diffère cependant de ce que Haidt (2003) appelle le DAAD (« distress at another's distress ») qui se réfère à la tendance pour un individu à être chagriné par le fait de voir ou d'entendre d'autres personnes émettre des signes de chagrin/détresse. Ce phénomène peut alors conduire les individus exposés à la souffrance d'une personne à fuir pour échapper à la détresse de cette personne au lieu d'aller l'aider.

1. Les émotions morales : un système d'outils au service de la décision

réaliser l'induction des émotions discrètes que l'on souhaite étudier ? Comment mesurer l'émergence de ces émotions chez les participants aux expériences ?

1.2.2.5. L'induction et la mesure spécifique des émotions morales

1.2.2.5.1. Les procédures d'induction

Pour étudier les affects et les émotions, les psychologues et les économistes ont mis au point différentes techniques afin de créer artificiellement en laboratoire, l'émergence d'un état affectif ou émotionnel spécifique. Il peut s'agir de techniques recourant au visionnage d'un film ou de photographies, capables de susciter l'état émotionnel recherché. Le passage distrayant/amusant d'un film peut par exemple mettre les participants de bonne humeur, permettant aux chercheurs de déterminer l'impact de la bonne humeur sur le comportement des participants dans un jeu économique. Cette technique peut être employée s'agissant des émotions morales en utilisant des passages de films ou des photos représentant des situations caractéristiques d'une émotion morale spécifique. Il est cependant nécessaire d'effectuer une étude préalable permettant de déterminer la capacité de ces matériels à réaliser l'induction émotionnelle souhaitée.

Moretti et di Pellegrino (2010) procèdent, par exemple, à l'induction de trois états émotionnels : l'état émotionnel de dégoût dont ils souhaitent comparer les effets avec l'état émotionnel de tristesse, et un état émotionnel neutre servant de base pour évaluer l'impact des deux états émotionnels précédents. Les auteurs choisissent de présenter une série de 13 photos aux sujets, prétendument pour une tâche liée à la mémoire, avant qu'ils ne jouent à un jeu de négociation. Ces 13 photos ont préalablement fait l'objet d'une expérimentation, visant à les sélectionner pour leur capacité à faire naître un état émotionnel de dégoût, de tristesse ou d'état émotionnel neutre, avec une même intensité, s'agissant des deux émotions à valence négative. L'observation des comportements réalisés par les individus dans le jeu, ainsi que la mesure de l'état émotionnel des sujets permettent de déterminer un éventuel impact du dégoût ou de la tristesse sur ces comportements.

Il peut également s'agir de techniques utilisant le rappel de souvenirs ou l'imagination. Dans le cas du rappel de souvenirs, les sujets sont amenés à se rappeler d'une situation dans laquelle ils ont ressenti un état émotionnel spécifique (celui qu'on souhaite étudier). Le souvenir de cette situation pouvant conduire les participants à ressentir à nouveau l'émotion ou l'état affectif en question, ou à imaginer la ressentir à nouveau. Dans le cas des procédures d'induction utilisant l'imagination, le plus souvent, les sujets doivent lire des vignettes décrivant une situation dans laquelle se trouve un personnage principal. Cette situation est choisie de manière à être susceptible de faire émerger le ou les états émotionnels spécifiques, dont on souhaite étudier l'impact sur les individus. Les sujets sont alors poussés à imaginer qu'ils vivent la situation décrite dans la vignette (dans le rôle du personnage principal), ce qui peut les conduire à ressentir ou à imaginer ressentir les émotions cibles. Cette technique, classiquement utilisée par les psychologues et les économistes, peut être employée dans le cadre des émotions morales.

Ketelaar et Au (2003), souhaitant analyser l'impact de la culpabilité dans une situation de coopération, proposent, par exemple, la procédure d'induction suivante. Lors de leur expérience, les sujets sont amenés à croire qu'ils jouent avec la même personne à un jeu du dilemme du prisonnier pendant 8 sessions de 10 répétitions chacune (soit 80 répétitions). En fait, ceux-ci jouent avec un ordinateur dont les comportements sont préprogrammés.

Après la quatrième série de 10 répétitions, les sujets doivent effectuer une tâche secondaire qui est, selon qu'ils sont dans la condition de contrôle ou dans la condition de culpabilité, de faire la description d'un jour normal ou de faire la description d'une situation où ils ont ressenti fortement de la culpabilité. L'impact de l'émergence de cette émotion est ensuite déterminé par l'observation du comportement des sujets, lors des 4 autres séries de 10 répétitions du jeu.

Une autre technique d'induction régulièrement employée par les chercheurs est de mettre directement les sujets dans une situation susceptible de faire émerger un état émotionnel spécifique. Les chercheurs employant cette technique ont dû être particulièrement inventifs pour pouvoir l'appliquer à l'étude des émotions morales. Celles-ci émergent, en effet, dans des conditions relativement spécifiques comme lorsque l'on est témoin, victime ou que l'on est soi-même responsable de la transgression d'une norme sociale ou morale.

Brown (1970) présente une expérience dans laquelle il est d'abord demandé aux individus de réaliser une tâche de perception sensorielle d'un objet, avant de réaliser une description des sensations ressenties pendant cette première tâche. Deux conditions sont présentes. Dans la condition contrôle, les sujets doivent toucher avec leurs mains une petite figurine en plastique pour en percevoir les contours (perception sensorielle), tandis que dans la condition « embarras », les sujets doivent sucer un objet en caoutchouc pour en percevoir les contours (perception sensorielle). On propose ensuite aux sujets plusieurs supports pour faire la description de leurs sensations pendant la première tâche. Les différents supports sont classés selon le niveau d'exposition des sujets vis-à-vis d'un panel d'individus censés évaluer la description des sujets. Du niveau d'exposition le plus élevé, au moins élevé, les sujets peuvent faire la description de la perception sensorielle de l'objet directement devant l'audience (le panel) en étant enregistré par vidéo ; faire un enregistrement vidéo de cette description, mais sans être obligé de se présenter devant le panel ; faire un enregistrement audio de cette description, sans avoir à se présenter devant le panel ; et faire cette description sur papier, sans signer de son nom et sans se présenter devant le panel. Les sujets sont rémunérés en fonction du niveau d'exposition du support choisi, plus le niveau d'exposition est élevé, mieux les individus sont payés (de 0 \$ à 1.50 \$). L'impact de l'état émotionnel d'embarras est donc observé à travers le choix du mode de description retenu pour retranscrire l'expérience sensorielle effectuée lors de la première tâche.

Ketelaar et Au (2003) présentent une deuxième expérience, lors de laquelle les sujets participent cette fois à un jeu de négociation. Ils jouent avec le même partenaire pendant deux répétitions du jeu qui s'effectuent à un délai d'une semaine d'intervalle. Afin d'analyser l'impact du sentiment de culpabilité sur le comportement des individus dans le jeu, les auteurs choisissent cette fois une procédure d'induction « naturelle » résultant directement de la situation de jeu dans laquelle se trouve les sujets et de leurs comportements dans ce jeu. Ils souhaitent, en effet, observer l'impact du sentiment de culpabilité chez l'offreur. Leur idée est de supposer que certains sujets peuvent ressentir un sentiment de culpabilité après avoir réalisé leur première offre, notamment les individus ayant réalisé des offres particulièrement égoïstes. Les auteurs peuvent ensuite évaluer les changements de comportements éventuels résultant de l'émergence de cette émotion, lors de la deuxième répétition du jeu.

1. Les émotions morales : un système d'outils au service de la décision

Plus récemment, DeSteno, Baumann, Barlett, Dickens et Williams (2010) ont utilisé une procédure d'induction originale pour étudier l'émotion de gratitude. Dans leur étude, les sujets participaient à un jeu similaire au dilemme du prisonnier (un jeu se jouant également à deux participants). Avant de participer au jeu, les sujets réalisent une première tâche, indépendante du jeu, en compagnie d'un partenaire. La procédure d'induction est réalisée à la fin de cette première tâche. Dans la condition « contrôle », la première tâche se déroule normalement tandis que dans la condition « gratitude », le sujet subit une difficulté d'ordre technique (manipulée par l'expérimentateur). Le partenaire (complice de l'expérimentateur) intervient alors en prodiguant de l'aide au sujet, dans le but de lui faire ressentir de la gratitude. L'impact de l'émotion de gratitude étant évalué, plus tard, lors de la participation des sujets au jeu économique, en comparant le comportement des individus de la condition « contrôle » et ceux de la condition « gratitude ».

Les procédés d'induction d'un état émotionnel sont donc le plus souvent adaptables à l'étude spécifique des émotions morales. Néanmoins, il est nécessaire de tenir compte de nombreux paramètres, comme les principaux éléments déclencheurs ou les tendances à l'action, lorsque l'on souhaite mettre en place une procédure d'induction concernant certaines émotions. C'est notamment le cas de la honte.

Hooge, Breugelmans et Zeelenberg (2008) indiquent que les études ayant tenté d'étudier l'impact que peut avoir le sentiment de honte sur le fait d'effectuer ou non des actions pro-sociales (comme l'étude de Hooge, Zeelenberg et Breugelmans, 2007) ont été limitées par le recours à des procédés d'induction inadaptés au sentiment particulier de honte. En effet, d'après les auteurs, il existe une distinction entre l'impact comportemental provenant de l'induction d'une émotion (d'un état émotionnel discret) s'effectuant de manière endogène d'une part et de manière exogène d'autre part. L'influence résultant de l'induction d'une émotion s'effectue de manière endogène lorsqu'elle concerne les comportements réalisés dans des situations en rapport avec l'évènement qui déclenche l'émotion (par exemple, l'influence que peut avoir le sentiment de peur à l'égard des animaux sur la décision d'aller ou non au zoo). A l'inverse, l'influence résultant de l'induction d'une émotion s'effectue de manière exogène lorsqu'elle concerne les comportements réalisés dans des situations sans rapport avec l'évènement qui déclenche l'émotion (par exemple, l'influence d'avoir vu un film drôle ou, au contraire, triste, sur le pourboire qu'on va laisser au taxi en rentrant du cinéma). D'après Hooge *et al.* (2008), cette distinction est pertinente et les auteurs considèrent qu'une des raisons pour lesquelles les précédentes études n'ont pas réussi à mettre à jour un effet fonctionnel positif du sentiment de honte, est que celles-ci ont étudié l'influence de l'état émotionnel de honte avec une induction exogène. Or, l'une des tendances à l'action de la honte est de faire son possible pour échapper à la situation responsable de l'émergence de ce sentiment. Il apparaît donc logique que peu ou pas d'effets soient dégagés lorsque l'on étudie l'influence de l'émergence de la honte sur les comportements des sujets dans le cadre de situations complètement étrangères à la situation d'induction, car du fait que ces situations (d'action ou de choix de comportements) sont étrangères à la situation d'induction, les individus ont déjà « échappé » à la situation responsable de la honte et ne sont plus sous l'influence de celle-ci. Sur cette base, les auteurs proposent d'étudier l'impact que pourrait avoir le sentiment de honte sur un processus décisionnel lorsque celui-ci est induit de manière endogène. Ainsi, Hooge *et al.* (2008) proposent une étude expérimentale utilisant une induction endogène du sentiment de honte.

De manière plus précise, les auteurs utilisent différents types d'induction de la honte, pour étudier l'influence de ce sentiment sur le comportement des agents dans un jeu de dilemme social. L'induction de la honte se fait à travers trois procédés, deux procédés impliquant une induction empathique de la honte (s'imaginer une situation, se rappeler d'une situation), ou par une induction « réelle » de la honte, à travers la performance des sujets dans une tâche préliminaire au jeu de dilemme social.

Les individus sont répartis entre la condition « contrôle » et la condition « honte ». Dans la condition « contrôle » la procédure d'induction est telle qu'elle entraîne un ressenti émotionnel neutre (les individus sont amenés à imaginer une situation neutre, à se rappeler d'une situation neutre, leur performance dans la tâche préalable est normale), tandis que dans la condition « honte », la procédure d'induction vise à faire émerger le sentiment de honte (les individus sont amenés à imaginer une situation dans laquelle ils ont ressenti de la honte, à se rappeler d'une situation honteuse, leur performance dans la tâche préalable est manipulée et rendue publiquement médiocre).

Dans chacune des deux conditions : « contrôle » et « honte », il existe également une distinction endogène versus exogène. A savoir que, les sujets sont amenés à s'imaginer qu'ils vont participer à un jeu de dilemme social avec un autre joueur. Ce joueur sera, dans la condition endogène, un individu en rapport avec l'induction de la honte (un individu présent dans la situation fictive ; un individu présent dans le souvenir, ou qui est au fait, de la situation ; un des participants lors de la tâche préliminaire), et dans la condition exogène, un individu sans rapport avec la procédure d'induction. L'impact du sentiment de honte est ensuite étudié à travers le comportement déclaré par les sujets dans le jeu de dilemme social.

Comme le montrent ces différentes illustrations, les procédés classiques d'induction d'un état émotionnel sont le plus souvent adaptables à l'étude spécifique des émotions morales. Néanmoins, l'exemple spécifique de l'étude de Hooge *et al.* (2008) témoigne de la difficulté de proposer une procédure d'induction qui soit pertinente vis-à-vis de l'émotion en question et de la difficulté d'utilisation de ces techniques. Il est ainsi nécessaire de tenir compte de nombreux paramètres, comme les principaux éléments déclencheurs ou les tendances à l'action, et d'introduire le plus souvent l'usage de la duperie, que ce soit pour justifier la présence de la procédure d'induction (pour ne pas dévoiler l'objectif de l'étude) ou lors de la mise en place de celle-ci (utilisation de complices, d'intelligences artificielles, etc.). De plus, il n'existe pas de consensus des spécialistes des émotions, vis-à-vis des conditions d'activation des différents états émotionnels spécifiques. Il est donc difficile d'être sûr que la procédure d'induction utilisée permet de déclencher l'état émotionnel que l'on souhaite étudier, ou que celui-ci est le seul à émerger. Par exemple, dans la technique du visionnage, un même extrait de vidéo dramatique peut susciter, selon les sujets, de multiples états émotionnels (tristesse, colère, culpabilité, honte). Afin d'évaluer le ressenti émotionnel des sujets, de pouvoir vérifier la bonne mise en œuvre des procédures d'induction, ainsi que de pouvoir proposer une analyse statistique de l'influence des émotions sur les comportements des individus, les chercheurs ont mis au point des techniques de mesure des émotions.

1.2.2.5.2. **La mesure des émotions**

Pour réaliser la mesure des émotions ressenties par les sujets participants lors des expérimentations, les psychologues et les économistes procèdent de trois façons différentes. Ils peuvent tout d'abord l'effectuer au travers de deux mesures objectives, l'une utilisant l'imagerie médicale qui va permettre d'étudier l'activation physiologique du corps (détection de zone activée dans le cerveau, mesure de la conduction de la peau, etc.) et l'autre se basant sur une étude des expressions corporelles ou physiologiques des individus (expression du visage, tonalité de la voix). Ils peuvent également passer par l'utilisation d'une mesure subjective de l'expérience émotionnelle s'effectuant par l'intermédiaire de l'auto-évaluation des sujets.

La première technique, basée sur l'utilisation des progrès et des outils de l'imagerie médicale, permet une mesure des émotions à travers une étude de l'activation de certaines zones du cerveau, responsables des émotions (pour une synthèse sur ce sujet, voir Camerer, Loewenstein et Prelec, 2005). Elle nécessite, cependant, un matériel médical assez lourd et onéreux à utiliser (comme l'Imagerie par Résonance Magnétique fonctionnelle ou IRMf) ce qui limite ses possibilités d'utilisation. De plus, la mesure spécifique des différentes émotions nécessite la connaissance de zones d'activités spécifiquement réservées à chacune d'elles. Or, des émotions comme la colère et le dégoût ont, par exemple, une zone d'activation commune (insula antérieure). Néanmoins, de nombreux espoirs sont suscités par les multiples découvertes réalisées dans le domaine de l'imagerie. Dans le futur, le développement de ce type de techniques pourrait déboucher sur une mesure objective plus performante des émotions des individus.

La mesure des émotions peut également s'effectuer grâce à l'étude des expressions faciales des individus ou à travers d'autres réponses de type physiologique (rythme cardiaque, pression sanguine, réponse électrodermale etc.). D'après Ekman (1982), il existe des émotions (ou émotions primaires) que les individus sont capables de différencier au travers des expressions du visage. Cependant, cette technique est peu adaptée à une mesure scientifique des émotions, elle se destine plutôt à une mesure des émotions entre individus. Certaines émotions peuvent, en effet, être feintes, ou imitées, de plus il est difficile de quantifier l'intensité du ressenti à partir des expressions faciales. Les expérimentateurs peuvent donc difficilement utiliser cette technique pour vérifier le bon fonctionnement de leur procédure d'induction.

La dernière technique de mesure est particulièrement adaptée à une utilisation scientifique de la mesure des émotions. Elle est, par ailleurs, celle qui est la plus utilisée dans les études de psychologie ou d'économie expérimentale du fait de son faible coût et de sa capacité à discriminer entre les émotions. Elle consiste en une mesure subjective de l'expérience émotionnelle du sujet (Frijda, 1986) et s'effectue par l'intermédiaire de questionnaires distribués au cours de l'expérience (après la procédure d'induction). Ces questionnaires sont conçus pour conduire les sujets à une auto-évaluation de la nature et de l'intensité de leurs émotions (présentes ou anticipées) sur une échelle de mesure ordinale. Cette technique permet donc d'obtenir des informations précises sur l'expérience émotionnelle des sujets (quelles émotions spécifiques le sujet a ressenti, avec quelle intensité, etc.) qui sont pertinentes vis-à-vis de l'étude spécifique des émotions. Cependant cette technique possède également des limites du fait de sa subjectivité ainsi que de la difficulté, pour les sujets, de pouvoir parfaitement comprendre et évaluer leur ressenti émotionnel.

1.2.2.6. Les principaux enseignements de la théorie des émotions morales

Au travers de notre analyse, certaines caractéristiques générales concernant les émotions morales se sont dégagées. Ces émotions sont, tout d'abord, en lien avec la notion de *norme sociale ou morale*. Les émotions morales à valence négative surviennent lorsque la violation d'une norme morale ou sociale est commise. A l'inverse, les émotions morales à valence positive émergent en parallèle avec le respect de ces normes, voire avec l'adoption de comportements qui dépassent le simple respect de celles-ci, des comportements de « beauté morale ».

Ces émotions ont un rôle particulier à jouer dans la manière dont les individus se comportent vis-à-vis de ces normes morales ou sociales. A travers leur tendance à l'action de *réciprocité négative* (vengeance, ostracisme, moqueries), la colère, l'indignation, le mépris ou le dégoût, poussent les individus à mettre en place des mécanismes de punition à l'encontre des transgresseurs. Ces punitions peuvent être matérielles à travers des actions et des prises de décisions ayant un impact (comme le fait de ne plus souhaiter faire affaire avec une personne, de porter plainte ou de rompre un contrat, etc.) ou psychologiques, parfois, à travers l'expression de ces émotions (expression violente de la colère, moqueries liées au mépris). Les émotions auto-conscientes négatives de honte, de culpabilité ou d'embarras peuvent alors survenir. Ces émotions, à forte valence négative, constituent une baisse de l'utilité pour les agents qui pèse dans leur processus de décision. En effet, l'anticipation des mesures punitives, résultant de la transgression d'une norme morale ou sociale, ainsi que l'anticipation de l'émergence d'émotions à valence fortement négative peut conduire à cadrer les comportements vers un respect de ces normes. Ceci a un intérêt particulier vis-à-vis de la capacité des agents à maintenir de bonnes relations interpersonnelles et à rester intégrés dans les différents cercles sociaux qui constituent leur réseau de relations. De plus, certaines émotions, comme la culpabilité ou l'embarras, peuvent conduire à des tendances à l'action qui visent à un rétablissement des relations mises en péril par une transgression (réparation du tort commis, actes conciliants).

Les émotions morales à valence positive jouent également un rôle dans le choix de comportement des agents, en permettant, à travers des mécanismes de récompense et de *réciprocité positive*, d'encourager et de favoriser le respect et le maintien des normes morales et sociales. L'anticipation joue à nouveau un rôle avec les émotions morales à valence positive. L'anticipation de l'émergence de la gratitude ou de l'admiration chez ceux qui nous entourent ainsi que du sentiment de fierté en nous, modifie également le processus de choix lorsqu'il est nécessaire d'adopter (ou non) une règle sociale ou morale. Ces émotions « positives » constituent une augmentation de l'utilité des individus et sont en lien direct avec leur bien-être. Elles encouragent également un dépassement positif (transgression positive) des normes morales, la gratitude, l'admiration ou la fierté pouvant conduire à l'adoption d'un comportement qui va au-delà et qui est « meilleur » que la norme. Les tendances à l'action des émotions de gratitude et d'admiration laissent enfin entrevoir la possibilité d'une capacité de ces émotions à se diffuser entre les agents, notamment à travers la réciprocité indirecte. Ce qui peut expliquer leur rôle primordial dans l'établissement des relations interpersonnelles comme le souligne Smith (1759) dès le 18^{ème} siècle.

En raison de leur impact sur le processus de décision des agents, de nombreux travaux analysent, dans le contexte du laboratoire ou en milieu naturel, l'influence que peuvent avoir les émotions morales dans le cadre des situations économiques d'interactions sociales.

1.3. Le rôle des émotions morales dans les interactions sociales

Les situations d'interaction sociale sont très présentes dans l'analyse économique dans la mesure où les relations d'échange sont, par essence, interpersonnelles. Il peut s'agir de situation de coopération, de négociation ou de confiance. Il existe plusieurs manières de les aborder dans l'analyse économique, qu'il s'agisse de la théorie des jeux, de la théorie des réseaux ou de l'étude des marchés. L'analyse de l'impact des émotions morales est pertinente dans ces situations, notamment dans le cadre des dilemmes sociaux.

Les dilemmes sociaux concernent des situations dans lesquelles la recherche de l'intérêt individuel s'oppose au bien-être collectif. Dans la littérature expérimentale, les dilemmes sociaux sont représentés via des jeux de *coopération* (dilemme du prisonnier, jeu du bien public, jeu de la ressource commune)⁹. Il s'agit de situations dans lesquelles, pour chaque agent, il existe une option opportuniste qui conduit toujours au gain le plus important dans le jeu, quel que soit le choix des autres. Dans ces jeux, le choix opportuniste de l'ensemble des joueurs mène à un gain inférieur (individuellement et collectivement) à celui qui serait obtenu si chaque joueur adoptait la solution coopérative (Hertel, Jochen Neuhof, Thomas Theuer and Norbert Kerr, 2000). C'est l'illustration du fameux problème du « passager clandestin ». Dans les dilemmes sociaux, les participants font face à un conflit entre optimiser leurs propres gains (dévier) et optimiser les gains du groupe (coopérer).

Dans le cadre de la *négociation*, le maintien de relations stables et saines implique le plus souvent des compromis et des sacrifices qui peuvent aller à l'encontre de l'intérêt individuel du négociateur. Lors d'une négociation, les différentes parties sont en conflit entre maximiser leurs propres gains et maximiser le surplus dégagé collectivement de la négociation. De plus, même si chacune des parties obtient un gain maximal dans la négociation lorsqu'il tente d'absorber le maximum possible du surplus, quelle que soit la décision des autres, le choix de cette stratégie par tous les joueurs peut conduire à une situation de rupture de la négociation entre les parties, impliquant des gains nuls voire négatifs si l'on prend en compte les coûts de recherche d'un partenaire. On obtient une solution pire (individuellement et collectivement) que si chacune des parties accepte de faire des compromis et décide de se montrer coopérative.

Dans le cadre de ces relations interpersonnelles spécifiques, le maintien ou le contournement des normes sociales ou morales et les émotions qui en résultent jouent un rôle primordial dans les choix et les comportements des agents. Ceux-ci sont en effet incités à transgresser certaines normes sociales ou morales – comme l'équité, la justice distributive, la réciprocité, la confiance ou la coopération – pour optimiser leurs intérêts. Cette transgression les expose à l'influence des émotions morales négatives (culpabilité, honte, mépris, colère) et aux mécanismes de punition qui y sont associés. Les émotions morales peuvent

9. Voir Eber et Willinger (2012) pour une revue.

ici décourager ou corriger des comportements opportunistes défavorables à l'intérêt collectif et favoriser, au contraire, les comportements prosociaux de coopération, d'altruisme ou de générosité. En parallèle, les émotions morales à valence positive induisent des mécanismes de récompenses qui favorisent le respect des normes, et l'adoption ou la diffusion de ces comportements prosociaux.

De nombreuses recherches en économie expérimentale et en psychologie sociale témoignent du rôle que jouent les émotions morales dans ce type de situations (Jourdheuil et Petit, 2015). Certaines de ces émotions ont, par ailleurs, été intégrées plus formellement dans la théorie économique, à travers leur incorporation dans des modèles dits de « préférences sociales », au sein desquels les agents économiques valorisent, non seulement leur propre bien-être, mais aussi celui des individus avec lesquels ils interagissent. Nous verrons, dans un premier temps, comment les émotions morales à valence négative favorisent l'émergence des comportements prosociaux, puis comment les émotions morales positives assurent la promotion et la pérennité de ces comportements.

1.3.1. Le rôle de la sanction morale affective dans l'émergence des comportements sociaux

1.3.1.1. Les effets prosociaux directs de la colère et du dégoût

1.3.1.1.1. La colère provoque la réaction de la victime de la transgression

L'un des premiers travaux marquant, en économie expérimentale, témoignant de l'impact des émotions morales à valence négative sur les comportements prosociaux est celui de Fehr et Gächter (2000). Les auteurs identifient un comportement punitif « altruiste » et montre dans un jeu du bien public¹⁰ répété, que l'existence d'une opportunité de punition coûteuse augmente significativement les niveaux de coopération, comparativement à la situation de référence, où l'opportunité de punition n'existe pas¹¹.

Pour pouvoir observer l'influence du mécanisme de punition, les auteurs font participer les sujets à 10 répétitions du jeu avec le mécanisme et 10 répétitions sans. De plus, les auteurs présentent deux conditions, une condition « partner », dans laquelle chaque groupe de 4 joueurs reste le même durant les 20 périodes et une condition « stranger »,

10. Dans ce jeu : 4 joueurs, ne pouvant pas communiquer, vont jouer ensemble. Chacun des joueurs reçoit une dotation initiale, 20 points expérimentaux, et peut soit les conserver, soit contribuer à la cagnotte commune (le bien public). Le gain de chaque joueur est égal au nombre de points expérimentaux qu'il décide de conserver, plus la totalité du montant investi dans la cagnotte par chacun des joueurs, multiplié par un certain coefficient, ici 0,4. Par exemple, si le premier joueur décide de conserver 10 points expérimentaux et que la cagnotte s'élève à 25 (sa contribution comprise), son gain à la fin du jeu est de 20 points expérimentaux ($10 + 25 \times 0,4 = 20$). La stratégie dominante pour chacun des joueurs est une contribution nulle, tandis que les paiements agrégés sont maximisés lorsque chaque joueur contribue à l'intégralité de sa dotation.

11. Dans les faits, cette possibilité de punition s'effectue de la façon suivante : après le premier stage de décision où chacun des joueurs décide du montant qu'il souhaite consacrer au bien public, chacun des joueurs a la possibilité, sur la base des contributions des autres joueurs de son groupe, de décider de les punir ou non. Chacun décide d'un montant de point de punition qu'il souhaite infliger aux autres joueurs. Le gain de chaque joueur étant diminué de 10 % de la somme totale des points de punition qui lui a été infligé par les autres joueurs. Cependant la punition a un coût et ce coût est strictement croissant avec le nombre de points que l'on souhaite infliger. Etant donné cette structure de jeu, si on se place d'un point de vue strictement maximisateur, un joueur s'en sort mieux lorsqu'il ne punit pas, puisque la punition représente un coût qui vient diminuer son gain dans le jeu.

1. Les émotions morales : un système d'outils au service de la décision

dans laquelle chaque groupe de quatre joueurs change à chaque période, de sorte qu'aucun des joueurs ne joue deux fois avec la même personne. Ainsi, dans la version « partner » il peut y avoir un intérêt stratégique à punir pour dissuader les joueurs opportunistes en vue des périodes futures. La condition « stranger » permet d'évacuer cet aspect stratégique. Pour autant, comme indiqué au-dessus dans les deux conditions (« partner » et « stranger »), lorsque les individus disposent d'un mécanisme de punition, les sujets utilisent cette possibilité et punissent les non contributeurs et les niveaux de coopération augmentent significativement. De plus, l'effet se produit également par anticipation, puisque l'on observe une augmentation significative du niveau de contribution au bien public dès la première période où la possibilité de punition apparaît (avant que le premier round de punition ne soit effectué). D'après Fehr et Gächter (2000), les passagers clandestins potentiels font face à une menace crédible qu'ils anticipent. Ceux qui coopèrent sont susceptibles de punir les passagers clandestins à la suite des très fortes émotions négatives ressenties durant le jeu (car ils se sentent exploités et qu'ils considèrent la situation comme injuste), en anticipation de cette sanction, les passagers clandestins potentiels vont augmenter leurs niveaux de coopération (Peeters et Vorsatz, 2013). Ainsi, d'après Fehr et Gächter (2000), la source principale de la punition (matérielle) est la colère, ou l'indignation, provoquée par le comportement transgressif et non coopératif de celui qu'on punit. Les auteurs montrent ainsi qu'avec la présence d'une punition coûteuse, réalisée sur la base de l'émotion de colère, une coopération quasi complète peut être obtenue et maintenue, bien que, selon l'hypothèse de rationalité et de recherche de l'intérêt individuel, il ne devrait y avoir aucune coopération.

Le mécanisme de punition permet également d'éviter la propagation des comportements de passagers clandestins. En effet, lorsqu'il est impossible de punir, ceux qui coopèrent n'ont aucune opportunité d'influencer le comportement des sujets égoïstes, ils sont donc incités à ajuster leurs contributions sur celles des passagers clandestins (Fehr et Gächter, 2000). Tandis que l'opportunité de punir leur permet de forcer les individus égoïstes à ajuster leurs niveaux de contribution vers les niveaux considérés comme appropriés. Enfin, la punition, coûteuse, est dite « *altruiste* » au sens où les contributeurs l'utilisent même lorsqu'ils savent qu'ils ne bénéficieront pas personnellement des effets positifs induits par la sanction, via la coopération (condition « stranger »). Ainsi, l'émotion de colère, du fait qu'elle entraîne un désir de vengeance et de réparation de l'injustice commise incite à prendre des mesures punitives, bien qu'elles soient coûteuses, contre les individus qui adoptent un comportement immoral. Ce mécanisme permet l'émergence de comportements prosociaux car il conduit à une coopération (et à un maintien de celle-ci) quasi complète. On voit donc clairement ici l'influence de l'émotion de colère (vertueuse, ou indignation) sur l'émergence de comportements prosociaux.

L'impact de la colère ou de la désapprobation sur l'émergence de la coopération a été généralisée dans d'autres travaux récents (Masclot, Noussair, Tucker et Villeval, 2003 ; Carpenter, Daniere et Takahashi, 2004). Masclot *et al.* (2003) observent que la coopération peut être soutenue par la simple désapprobation morale du comportement observé des autres joueurs (punitions non monétaires). Carpenter *et al.* (2004) montrent, qu'en milieu naturel, de nombreux participants sont prêts à exprimer leur désapprobation (au moyen d'une image représentant un visage mécontent) à ceux qui se montrent non coopératifs.

Des résultats similaires sont également obtenus par Xiao et Houser (2005), Bosman et van Winden (2002) et Ben-Shakhar, Bornstein, Hopfensitz et van Winden (2007) dans des jeux de négociation¹². Ces travaux indiquent le rôle dissuasif et efficace de la colère ou du mépris lorsque l'un des partenaires fait une proposition d'offre perçue comme injuste et intentionnelle (Blount, 1995). Xiao et Houser (2005) étudient le rôle des émotions négatives, dont la colère, dans des situations de négociation. Les auteurs montrent que, lorsque les individus n'ont pas d'autres moyens de communication, ils choisissent d'exprimer leurs émotions négatives, vis-à-vis d'une offre injuste dans le jeu de l'ultimatum, à travers une punition coûteuse, via le refus de l'offre entraînant un gain nul pour les deux joueurs. On voit ici que les individus sont à nouveau prêts à faire des sacrifices, sur la base des émotions négatives ressenties, pour pouvoir punir les sujets non coopératifs (qui font des offres considérées comme injustes). Ben-Shakhar *et al.* (2007) apportent, par ailleurs, de nouvelles preuves empiriques du rôle prépondérant de la colère et du mépris dans la décision de punir à travers la destruction coûteuse dans un jeu de négociation à un coup. Pour autant, Xiao et Houser (2005) montrent également que lorsque les sujets ont l'opportunité d'exprimer leurs émotions de manière explicite (sur papier), ils utilisent significativement moins la punition coûteuse (rejet de l'offre).

Le message écrit peut alors être interprété comme une punition psychologique qui est susceptible de faire naître des émotions comme la honte ou la culpabilité chez l'offreur injuste. Et lorsque les sujets ont l'opportunité de pouvoir s'exprimer à travers un message écrit, ils sont face à un choix entre deux types de punition, l'une coûteuse et qui sanctionne de manière monétaire, l'autre non coûteuse et qui sanctionne de manière non monétaire. Ainsi les individus peuvent être prêts à faire des sacrifices monétaires pour exprimer leur mécontentement, mais peuvent préférer d'autres types de punitions, moins ou non coûteux, s'ils en ont l'opportunité. La simple expression d'une émotion négative comme la colère peut avoir une influence sur le comportement des sujets. En effet, dans une étude de 2009, Xiao et Houser présentent une autre expérience en négociation, utilisant cette fois le jeu du dictateur, dans laquelle ils observent significativement moins de comportements maximisateurs de la part des dictateurs, lorsque ceux-ci peuvent recevoir un message écrit de la part du récipiendaire après le dévoilement de leur offre (comparativement à la version standard du jeu). Leurs résultats montrent ainsi que l'anticipation de la punition psychologique contenue dans le message suffit pour faire émerger des comportements prosociaux. Même si les auteurs rappellent que les menaces de sanctions monétaires (présentes dans le jeu de l'ultimatum), bien qu'inefficaces économiquement lorsqu'elles sont réalisées (les deux joueurs se retrouvant alors avec un gain nul), sont plus performantes pour dissuader les comportements égoïstes.

Moretti et di Pellegrino (2010) retrouvent ce type de mécanisme en s'intéressant à l'émotion de dégoût. Leur étude montre, en particulier, la proximité entre la colère et le dégoût, deux émotions qui sont activées dans des zones neuronales très proches. Un travail de Sanfey, Rilling, Aronson, Nystrom, et Cohen (2003) avait montré par ailleurs, à l'aide d'instruments d'imagerie médicale, que le rejet d'offres faibles dans le jeu de l'ultimatum est corrélé avec une activité importante dans une zone du cerveau reliée à l'émotion de colère. Moretti *et al.* (2010) évaluent l'impact de l'induction de l'émotion de

12. Les jeux de négociation sont ceux de l'ultimatum ou du « pouvoir de prendre » (« power-to-take game »), dans lesquels l'un des joueurs fait une offre que le second peut accepter ou refuser, ou encore celui du « dictateur » dans lequel le receveur ne dispose plus de droit de veto (pour une revue sur ces jeux, voir, respectivement, Petit (2010) et Petit et Rouillon (2010)).

dégoût, comparativement à l'induction de la tristesse ou à une induction neutre, sur le comportement des répondants dans un jeu de l'ultimatum. Ils expliquent que certaines des zones activées dans le cerveau étant les mêmes, lorsque les individus ressentent du dégoût à la suite d'un stimulus olfactif ou visuel, et à la suite de la réception d'une offre injuste dans le jeu de l'ultimatum (d'après des mesures effectuées à l'aide d'IRM), la violation de la norme d'équité et le dégoût viscéral (à la suite d'un stimulus sensoriel), peuvent impliquer des mécanismes neuraux qui se recoupent. Ils supposent ainsi, du fait que l'émotion de dégoût est congruente au processus affectif qui s'effectue lorsque l'on est confronté à une offre inéquitable dans un jeu de l'ultimatum, que l'induction du dégoût chez des joueurs répondant, devrait entraîner plus de refus de leur part, vis-à-vis d'offres inéquitables, que l'induction de la tristesse ou qu'une induction neutre.

D'après les résultats, les individus de la condition dégoût ont un taux d'acceptation de l'offre, dans le jeu de l'ultimatum, significativement plus faible que dans les deux autres conditions. Cette différence provient du fait que les individus préalablement dégoutés (par la procédure d'induction) rejettent, de manière significativement plus importante, les offres faibles (< 50 % de la dotation initiale de l'offreur), que les sujets des deux autres conditions. Ces résultats montrent comment le sentiment de dégoût, ici exacerbé par l'induction effectuée avant la participation au jeu, participe, de manière similaire à l'émotion de colère, à la décision de rejet des individus vis-à-vis des offres faibles ou injustes dans le jeu de l'ultimatum, et donc à la décision de réaliser un acte punitif à l'égard de l'offreur vis-à-vis de son comportement transgressif. Dans le cadre d'interactions répétées, ce mécanisme punitif peut, comme dans le cas des études développées précédemment, aboutir à l'adoption de comportements socialement ou moralement adaptés de la part des transgresseurs punis.

1.3.1.1.2. Sur un tiers

La réaction de la victime d'une transgression, via la punition, peut ne pas être suffisante pour empêcher l'émergence de comportements opportunistes. Lorsque, par exemple, un individu envisage de voler quelqu'un, s'il sait que d'autres personnes que la victime (des témoins des membres de sa communauté ou encore la société) peuvent également entreprendre des mesures punitives contre lui, l'anticipation de ces mesures sera probablement plus efficace, pour dissuader le voleur potentiel. Dans ce cas, le maintien des normes sociales est obtenu via la participation d'un tiers, une personne qui ne subit pas le comportement immoral d'autrui mais qui en est le témoin.

En outre, si le désir de remédier, par vengeance, à une situation injuste est présent lorsque la personne qui ressent la colère est elle-même la victime de l'injustice, celui-ci peut également intervenir lorsque cette injustice touche une tierce personne. Dans ce cas, la moralité de l'acte de punition est évidente, puisqu'il correspond à un renforcement des normes sociales ou morales de la part de personnes outragées (avec un fort sentiment de colère) mais qui ne sont pas directement affectées par la violation qui a été commise. D'après Fehr et Fischbacher (2004a), ce type de comportement est primordial, puisque, selon eux, la plupart des normes ne seraient pas maintenues si seules les personnes directement concernées imposaient des sanctions.

Un article récent de Carpenter et Matthews (2010) apporte des preuves empiriques de l'existence et de la pertinence de ce type de punition. Les sujets participent à un jeu du bien public similaire à celui présenté par Fehr et Gächter (2000). Lors de chaque session de

l'expérience, 2 groupes de 4 joueurs jouent simultanément et de manière distincte (chacun de son côté) s'agissant de la contribution au bien public, mais les membres de chacun des groupes peuvent interagir (ou non selon la condition), avec les membres de l'autre groupe, lors de la phase de punition (coûteuse, comme dans Fehr et Gächter, 2000). En particulier, certaines conditions, autorisent les sujets d'un groupe à punir les sujets de l'autre groupe, vis-à-vis de leur niveau de contribution. Cette punition, si elle est réalisée, ne provient pas d'un désir de vengeance, en lien avec le fait de subir un comportement transgressif et non coopératif, puisque les gains obtenus par les joueurs dans un groupe, ne sont aucunement affectés par le comportement contributif des joueurs de l'autre groupe. Celle-ci peut alors provenir de l'émergence d'un sentiment d'indignation résultant de l'observation d'un comportement de passager clandestin.

Les auteurs montrent que les niveaux de punition sont plus élevés lorsque les sujets punissent les individus au sein de leur groupe que lorsqu'ils décident de punir ceux de l'autre groupe, cependant, d'après leurs données, dans la condition où les sujets peuvent seulement punir les individus de leur propre groupe, cela ne suffit pas pour entraîner une augmentation significative du niveau de contribution par rapport à la condition de contrôle (où il n'y a pas de possibilité de punition). L'introduction d'une possibilité de punition par une tierce personne (non impliquée) permet quant à elle d'obtenir cette augmentation significative et de dissuader les comportements déviants. Et même si, lorsque seuls les membres d'un des deux groupes ont la possibilité de punir les membres de l'autre groupe, 10 % seulement, des sujets ont procédé à une « punition altruiste » (comparativement aux 60 % obtenu dans l'expérimentation de Fehr et Fischbacher (2004a) qui utilise un jeu du dictateur), L'indignation joue tout de même un rôle (significatif) puisque, la punition étant coûteuse, on aurait pu s'attendre à ce que le niveau de « punition altruiste » soit nul.

En déclenchant des représailles, la colère et le dégoût ont un impact direct sur le comportement de la victime ou du tiers. Ces émotions sont ici l'expression d'une réciprocité négative formalisée par Rabin (1993). Ce qui leur confère un rôle important dans la régulation de la prise de décision, du fait de leur capacité à dissuader les comportements transgressifs. Ces émotions de condamnation d'autrui peuvent également déclencher d'autres processus émotionnels chez le transgresseur, qui l'inciteront à ne pas enfreindre la règle sociale en cours. C'est le cas notamment lorsque celui qui est le sujet moral d'une juste indignation, ressent de la culpabilité, des remords ou de la honte, qui l'obligent à reconsidérer son comportement initial. Le maintien de la norme provient ici à la fois d'une régulation externe (la victime ou la personne tierce) et d'une régulation interne ou intériorisée, via une émotion auto-consciente.

Une étude récente de Reuben et van Winden (2010) apportent des preuves de l'existence de ce type de phénomène dans les situations de dilemmes sociaux. Les auteurs présentent un jeu du « Prendre ou à laisser » durant lequel les sujets participent à deux répétitions du jeu. Pour limiter les biais d'expérimentation, le rôle de chacun des sujets est déterminé aléatoirement à chaque répétition, ainsi un individu peut être amené à changer de rôle entre les deux répétitions. Toutefois, les sujets qui intéressent les auteurs sont ceux qui ont le rôle d'offreur pendant les deux sessions du jeu pour pouvoir déterminer ce qui motive leur comportement. Entre les deux sessions, les offreurs doivent remplir des questionnaires à propos des émotions qu'ils ont ressenties à la suite de la première session, ils doivent également indiquer leur perception de ce que serait un comportement juste dans leur si-

1. Les émotions morales : un système d'outils au service de la décision

tuation (quel pourcentage de la dotation initiale du répondant doit-on prélever pour être juste?).

Ces informations ont permis à Reuben *et al.* (2010), de déterminer que les individus qui décident de diminuer le pourcentage qu'ils prélèvent sur la dotation de l'autre joueur, après avoir subi une punition lors de la première répétition du jeu, reportent de plus fortes intensités de ressenti des émotions de culpabilité et de honte. On peut donc décomposer la séquence de comportements de la manière suivante. L'offreur choisit un taux de prélèvement sur la dotation de l'autre joueur. Suivant le taux de partage qui est choisi, celui-ci peut constituer une violation de la norme d'équité (au moins du point de vue du répondant), auquel cas, sur la base des très fortes émotions de colère ou de mépris, l'autre joueur peut entreprendre des mesures punitives en détruisant une partie de son capital (ce qui aura pour effet de diminuer ce que l'offreur prélève au répondant) (Bosman et van Winden, 2002). A la suite de cette punition, les émotions de culpabilité ou de honte peuvent émerger chez l'offreur, et inciter celui-ci à modifier ses comportements vers une diminution du taux de prélèvement lors de la deuxième répétition, c'est-à-dire vers un comportement plus respectueux de la norme d'équité.

On retrouve l'enchaînement de cette séquence dans l'étude de Joffily, Masclet, Noussair, et Villeval (2014), qui montre expérimentalement, dans le cadre du jeu du bien public, comment les décisions individuelles sont capables de déclencher l'émergence des émotions et comment celles-ci peuvent, en retour, modifier les comportements en faveur de la coopération. Dans leur expérience, l'activité électrodermale des participants est mesurée pendant le jeu, afin de déterminer l'ampleur de leurs réactions émotionnelles face à l'interaction qui est cours et, s'agissant des sujets de la condition « punition », les résultats montrent que les individus qui coopèrent le plus ressentent des émotions à valence négative (probablement la colère ou la déception) lorsqu'ils apprennent que certains membres de leur groupe n'ont pas respecté la norme de coopération. Chez ces individus, le fait de punir les agents se comportant en « passager clandestin » est également à l'origine d'un ressenti émotionnel, de valence positive, ce qui fait dire aux auteurs que l'action d'infliger une sanction permet de réduire la désutilité liée au mauvais comportement d'autrui. Par ailleurs, les sujets les moins coopératifs ressentent des émotions à valence négative (culpabilité, honte, peur etc.) lorsqu'ils sont informés de la faiblesse relative de leur contribution, ainsi que lorsqu'ils reçoivent une punition de la part d'un partenaire. De plus, lorsqu'ils sont punis, ces derniers ont tendance à augmenter leur niveau de contribution lors de la répétition suivante du jeu, ce qui conduit les groupes à converger vers une norme de contribution permettant un niveau conséquent de coopération.

Ces résultats témoignent donc de l'impact conjoint des deux catégories d'émotions morales à valence négative dans l'émergence d'un comportement (plus) prosocial dans une situation de dilemme social. La punition matérielle fonctionne ici, main dans la main, avec l'émergence des émotions auto-conscientes à valence négative pour les agents transgressifs vers l'adoption de comportements prosociaux. Cette étude témoigne du rôle important, de l'anticipation de l'émergence de ces émotions auto-conscientes dans le processus de décision des agents.

1.3.1.2. **Les effets prosociaux des émotions anticipées : culpabilité, honte et embarras**

D'après Frank (1988), les émotions morales auto-conscientes à valence négative incitent les individus à choisir les comportements prosociaux qui possèdent un intérêt stratégique à long terme. Elles s'opposent à l'intérêt personnel, davantage orienté vers les récompenses immédiates (défection, trahison, procrastination) en faisant miroiter les conséquences émotionnelles négatives de comportements opportunistes. De tels agissements suscitent en effet des émotions déplaisantes (culpabilité, honte, embarras) qui peuvent être anticipées et rationnellement intégrées par l'individu dans sa fonction d'utilité de long terme. En régulant à long terme les comportements sociaux, ces émotions ont ainsi un rôle crucial dans la construction et le maintien des relations économiques interpersonnelles.

L'expérimentation de Brown (1970) montre très bien le mécanisme d'anticipation à l'œuvre en présence d'interaction sociale. Dans son étude, les participants ont le choix de plusieurs supports médias pour présenter la description d'une tâche de perception sensorielle qui peut être, selon la condition à laquelle ils appartiennent, embarrassante ou neutre. La rémunération perçue pour cette description dépend positivement du niveau d'exposition du média choisie. Classiquement, un individu rationnel doit se porter vers le support où le niveau d'exposition est le plus élevé pour faire sa description. Or, les résultats montrent que par rapport au paiement maximal de 1.50 \$, les individus de la condition « embarras » sacrifient significativement plus d'argent (0.63 \$ en moyenne) que les individus de la condition contrôle (0.33 \$ en moyenne). Un des résultats importants de cet article est qu'il montre que les individus sont prêts à faire des sacrifices financiers pour ne pas ternir l'image de soi qu'ils donnent aux autres, au travers d'une situation publiquement embarrassante. Ainsi les individus peuvent être prêts à faire des sacrifices en choisissant d'adopter un comportement social ou moral (coopérer, ne pas agir en passager clandestin) plutôt qu'un comportement maximisateur, dans le but de ne pas se retrouver dans une situation embarrassante et donc d'éviter de ressentir le sentiment d'embarras.

Les travaux plus récents de Ketelaar et Au (2003) valident empiriquement l'argument donné par Frank (1988). Les auteurs s'intéressent au rôle de l'émotion de culpabilité sur le comportement des agents dans deux jeux de dilemmes sociaux. Lors de leur première expérience, les sujets jouent de manière répétée au dilemme du prisonnier et subissent une induction de la culpabilité recourant à leur mémoire. Ketelaar et Au (2003) montrent alors que les sujets de la condition « culpabilité » (devant se souvenir d'un événement où ils ont ressenti cette émotion) adoptent un comportement significativement plus coopératif que les sujets de la condition « contrôle » (devant faire la description d'une journée normale) lors des répétitions du jeu suivant la procédure d'induction. De plus, l'effet de l'induction de la culpabilité est significativement plus important chez les sujets ayant eu un comportement non coopératif lors des répétitions effectuées avant l'induction. Ainsi, les auteurs montrent que l'induction du sentiment de culpabilité peut motiver l'émergence de comportements prosociaux coopératifs, en particulier lorsque cette induction est congruente avec les comportements précédemment adoptés par les sujets (comme le fait d'avoir dévié lors des précédentes répétitions, comportement pour lesquels les sujets peuvent s'en vouloir).

1. Les émotions morales : un système d'outils au service de la décision

Lors de la deuxième expérience, concernant cette fois un jeu de l'ultimatum. Les auteurs montrent que la vaste majorité des sujets ayant rapporté un sentiment de culpabilité après avoir effectué une première offre égoïste (proposant un partage où ils obtiennent plus que leur partenaire), changent de stratégie une semaine après et font une deuxième offre plus généreuse lors de la deuxième répétition du jeu. A l'inverse, ceux qui ont réalisé une première offre égoïste mais qui n'ont pas ressenti de sentiment de culpabilité, n'ont pas modifié leur comportement d'offre à la deuxième répétition. Les auteurs montrent à nouveau que l'émergence du sentiment de culpabilité, cette fois résultant du comportement même des sujets dans le jeu, permet d'orienter leur comportement vers davantage de générosité.

Le sentiment de culpabilité pourrait ainsi servir d'information vis-à-vis du coût à long terme (punitions ou dégradation voire exclusion des relations sociales futures) que représente le fait d'agir de manière non coopérative ou égoïste (comportements violant les normes morales d'équité/justice), en succombant toujours à la tentation immédiate de transgresser les normes sociales ou morales en vigueur, dans le but de satisfaire son intérêt individuel. La douleur psychique subie lorsque l'émotion survient, fournit donc une information sur ce coût futur et incite les individus à modifier leurs comportements vers des comportements plus respectueux des normes auxquelles ils font face. L'émotion de culpabilité peut donc être interprétée comme une auto-punition, effective dès l'accomplissement de l'acte, signalant à l'individu les effets néfastes que peuvent avoir ses comportements déviants sur son bien-être futur. Cette information lui étant utile, au moment où il doit effectuer un nouveau choix de comportement, durant lequel, l'anticipation de l'émergence du sentiment de culpabilité ou des émotions de colère ou de dégoût (conduisant à des mesures punitives de la part des autres individus) peut influencer sa prise de décision finale. Ce signal peut alors être intégré dans la fonction d'utilité de long terme de l'individu.

L'impact de la honte en tant que signal permettant de mobiliser à long terme des comportements prosociaux est moins évident. Il existe effectivement une abondante littérature expérimentale montrant que les individus ont tendance à adopter des comportements altruistes lorsqu'ils sont observés publiquement, ces comportements étant beaucoup plus limités dans la sphère privée (pour une revue, voir Petit et Rouillon, 2010). Ceci soutient que le comportement de don serait davantage issu d'un désir d'apparaître comme équitable plutôt que d'une véritable préférence pour l'équité. Certains travaux théoriques en ont déduit que le sentiment de honte, associé au désir de ne pas apparaître comme égoïste, pouvait motiver le comportement de don (Neilson, 2009 ; Dessi et Zhao, 2011). En particulier, Dana, Cain et Dawes (2006), présentent un jeu du dictateur avec option de sortie. Cette spécificité implique que le joueur dictateur, après avoir effectué le choix de répartition de sa dotation initiale (10 €) entre lui-même et le récipiendaire, se voit offrir le choix suivant : laisser le jeu continuer et obtenir les gains correspondants à sa décision de répartition, auquel cas le récipiendaire obtient lui aussi son gain dans le jeu et apprend de fait la décision prise par le dictateur avec lequel il joue, ou bien prendre l'option de sortie. C'est-à-dire obtenir 9 €, le récipiendaire n'obtenant rien du tout, mais, en contrepartie, s'assurer que le récipiendaire ne sera pas informé du jeu et de ses règles ni du comportement du dictateur dans le jeu. Théoriquement, on pourrait penser qu'un dictateur ne devrait jamais choisir l'option de sortie. Soit il est animé par des considérations purement égoïstes, auquel cas il vaut mieux pour lui participer au jeu et obtenir le gain maximal, supérieur à l'option de sortie. Soit il est animé par des considérations

altruistes, qui le poussent à vouloir prendre soin du récipiendaire en lui laissant une partie de la dotation initiale, auquel cas il est également plus pertinent pour lui de rester dans le jeu puisque s'il prend l'option de sortie, le récipiendaire n'obtient rien.

Pourtant, d'après Dana *et al.* (2006), 40 % des sujets ayant fait une offre positive choisissent l'option de sortie. Les auteurs expliquent que l'anticipation de la honte, liée au fait de ne pas donner, pourrait donc être à l'origine du don positif, qui peut être assimilé à un comportement prosocial visant à l'adhésion de la norme d'équité, chez ces personnes. L'effet de la honte pouvant être évité dans le cadre de l'option de sortie, ils choisissent cette option qui leur permet de maximiser leurs gains dans le jeu sans subir les effets négatifs de la honte. On voit ainsi comment le fait de vouloir éviter le ressenti négatif de la honte, issu de l'adoption d'un comportement égoïste, peut conduire à l'adoption de comportements moralement ou socialement adaptés. Ces comportements ne sont, cependant, réalisés que lorsque les individus n'ont pas la possibilité de s'échapper à travers le choix de l'option de sortie. Pour autant, les résultats obtenus par Dana *et al.* (2006), ne ciblent pas explicitement l'influence du sentiment de honte sur le comportement des sujets.

Sur le plan expérimental, les effets de la honte induite sur les comportements sociaux sont mitigés. Hooge *et al.* (2007), utilisant un protocole similaire à celui de Ketelaar *et al.* (2003), retrouvent un effet significatif de la culpabilité, mais aucun impact de la honte. La honte serait bien ainsi cette émotion « détestable et diabolique » qui, mettant en doute l'estime de soi de l'individu, l'incite à fuir la relation sociale plutôt qu'à la restaurer, comme est susceptible de le faire la culpabilité. Une étude plus récente de Hooge *et al.* (2008) remet en question cette vision univoque. Les auteurs partent de l'idée, admise chez la plupart des théoriciens, que les émotions, et notamment les émotions morales, sont fonctionnelles, dans le sens où elles promeuvent des comportements qui ont des conséquences bénéfiques pour l'individu ou la communauté (Fridja, 1986). Ils se demandent alors, comment une émotion telle que la honte, qui est une émotion très douloureuse, pourrait être une émotion fonctionnelle, si elle n'a que des conséquences négatives ?

A travers la mise en place de procédés d'induction classiques, incluant des procédures endogènes, ils étudient l'influence de ce sentiment sur le comportement des agents dans un jeu de dilemme social (« give-some dilemma »). Un autre aspect qui intéresse les auteurs est celui de la personnalité des individus, ils réalisent donc également la distinction, dans leur analyse, entre deux catégories de sujets, les prosociaux et les pro-selfs. Ils utilisent pour cela un test de mesure de l'orientation sociale. Ils justifient cette procédure du fait que les individus prosociaux agissent de manière prosociale de façon régulière et l'influence du sentiment de honte devrait donc s'effectuer en priorité sur les individus pro-selfs, qui agissent le plus souvent dans leur propre intérêt.

Les résultats obtenus correspondent à ce que les auteurs supposaient, à savoir que l'on peut dégager une influence significative de l'induction de l'état émotionnel (discret) de honte sur le comportement prosocial des sujets dans le jeu. Cependant, cet impact n'est significatif que lorsque l'on considère les individus catégorisés comme pro-selfs et surtout lorsque l'induction du sentiment de honte est endogène à la situation dans laquelle le comportement des sujets est étudié, c'est-à-dire lorsque le comportement d'aide ou de coopération est destiné à celui qui a été témoin de la situation honteuse. Ainsi, Hooge *et al.* (2008) démontrent avec ce travail que le sentiment de honte, possède un aspect fonction-

1. Les émotions morales : un système d'outils au service de la décision

nel et permet, au même titre que l'émotion de culpabilité, l'émergence de comportements prosociaux dans certaines conditions.

Si on rapproche cette étude avec celle de Ketelaar et Au (2003), on se rend compte que l'impact de ces deux émotions s'effectue de manière significativement différente, selon qu'on se comporte de manière égoïste ou généreuse pour Ketelaar et Au (2003), ou selon qu'on est un pro-self ou un prosocial pour Hooge *et al.* (2008). Ces résultats similaires tendent à montrer que l'influence des émotions s'effectue de manière différente selon la personnalité des individus. En outre, de manière semblable au fait que dans l'expérience de Ketelaar et Au (2003), tous les participants ayant agi de manière égoïste (selon les critères des auteurs) lors de la première répétition du jeu (de l'ultimatum) n'ont pas tous déclaré avoir ressenti de la culpabilité, les sujets de l'expérimentation de Reuben *et al.* (2010), ayant subi une punition lors de la première répétition du jeu, n'indiquent pas tous avoir ressenti de la honte ou de la culpabilité. La différence d'influence de ces émotions, peut donc s'expliquer par une différence dans la capacité des individus à les ressentir. Alors, qu'est ce qui explique ces différences de ressenti émotionnel ? Ketelaar et Au (2003) proposent plusieurs explications. Tout d'abord, ils considèrent que cela peut faire partie de la personnalité des individus. Certains sont généreux de nature, agissent de manière prosociale et ne ressentent donc pas de culpabilité, certains peuvent initialement se comporter en égoïste, mais tenter de se rattraper après avoir ressenti de la culpabilité, et d'autres, être égoïste de nature et ne pas ressentir cette émotion ou encore ne pas en subir l'influence. Ainsi, la personnalité des individus pourraient conditionner leur capacité ou leur tendance, à ressentir et à subir l'influence de certaines émotions comme la honte ou la culpabilité. Cet aspect pourrait être un point important s'agissant de la modélisation du processus de décision des agents, lorsque l'on souhaite y intégrer le rôle des émotions.

Ketelaar et Au (2003) expliquent également que, étant donné que les participants jouent avec des camarades de classes sans procédure d'anonymat, les différences de ressenti de culpabilité peuvent provenir des différences de distances sociales existantes entre les individus (jouer avec son meilleur camarade de cours, ou jouer avec une vague connaissance assise à côté de nous au moment de choisir son partenaire dans le jeu). L'existence de liens sociaux a donc également vocation à jouer sur la façon dont les processus émotionnels modifient les choix individuels (van Winden, Stallen, Ridderinkhof, 2008).

Reuben *et al.* (2010), apportent une explication complémentaire. En effet, grâce à leurs questionnaires, ils ont pu déterminer que les joueurs ayant choisi un taux de prélèvement qu'ils considèrent comme injuste, ressentent avec plus d'intensité les émotions de honte ou de culpabilité. Les différences de niveau de ressenti pourraient également provenir des différences de croyances vis-à-vis de ce qu'est la norme. Ceci pourrait expliquer pourquoi le fait de laisser les individus communiquer avant de participer (ou pendant la participation) à un jeu de dilemme social, permet souvent d'obtenir des comportements plus coopératifs, du fait que cette procédure peut permettre de coordonner les croyances des différents participants sur le niveau auquel se fixe la norme d'équité dans le jeu et d'agir ensuite en conséquence à travers leur comportements. Ces résultats laissent à penser que l'éducation reçue par les agents peut être un facteur important dans leur capacité à ressentir et à être influencés par les émotions morales. En effet, les croyances vis-à-vis des normes morales ou sociales des individus sont en partie déterminées par l'éducation qu'ils reçoivent.

En conclusion, les deux familles d'émotions morales à valence négative peuvent donc avoir conjointement un impact sur le comportement et le processus de décision des agents dans le cadre des situations de dilemmes sociaux. En particuliers par des mécanismes punitifs, issus directement du ressenti émotionnel (émotions auto-conscientes) ou entraînés par celui-ci (émotion de condamnation d'autrui). Néanmoins, ces émotions ont certaines limites.

1.3.1.3. **Les effets non désirés des émotions morales à valence négative**

Les émotions morales ont des effets prosociaux bénéfiques directs, via les comportements de réciprocité négative de leurs émetteurs, et indirects, via l'anticipation d'émotions auto-conscientes chez celui qui viole la norme sociale. Ces émotions ont néanmoins certaines limites qui ont également été identifiées expérimentalement :

1. La colère, l'indignation, le mépris ou le dégoût peuvent tout d'abord se révéler contreproductives en suscitant à leur tour des émotions négatives chez l'individu qui a enfreint la norme collective. Ceci implique un risque de représailles, d'escalade du conflit et de blocage des négociations (Pietroni, van Kleef, De Dreu et Pagliaro, 2008, Hopfensitz et Reuben, 2009 ; voir Petit, 2010, pour une revue) ou bien la dégradation des relations interpersonnelles entre les agents au sein d'un groupe ou d'une entreprise. Par exemple, des recherches ont mis en évidence les effets négatifs de la punition, dont on attribue la cause à l'émergence d'émotions à valence négative, sur les relations interpersonnelles, en ce sens que les individus qui se savent sous la menace d'une punition éventuelle, au moment d'une décision économique, ont tendance à faire leur choix de manière plus rationnelle et en étant moins concerné par le bien-être d'autrui (Xiao, Houser, et Montague, 2009).
2. Les mécanismes de punitions engendrés par ces émotions représentent par ailleurs un coût économique non négligeable pour ceux qui les mettent en œuvre. Il en résulte qu'en termes de bien-être social la mise en œuvre d'une procédure de sanction n'a pas toujours l'effet escompté même lorsqu'elle rectifie les comportements opportunistes.
3. Les effets de la sanction ne sont en général pas permanents. Dans l'expérience de Fehr et Gächter (2000), le comportement égoïste des joueurs réapparaît massivement et spontanément dès l'élimination de la menace de sanction. Dans le jeu du bien public, un comportement bien identifié de « fin de jeu » – la chute drastique des contributions à l'annonce de la fin du jeu – révèle également cet aspect stratégique. Dans le jeu du dictateur, les effets potentiels de la honte sur la capacité de don sont éliminés lorsque l'on propose aux participants une option de sortie leur permettant de protéger leur intérêt individuel à l'insu du récipiendaire (Dana, Cain et Dawes, 2006).
4. Par définition, les émotions à valence négative représentent une perte en termes d'utilité pour ceux qui les ressentent. En conséquence, elles nuisent au bien-être individuel. Certains travaux en psychologie soulignent en particulier la corrélation né-

1. Les émotions morales : un système d'outils au service de la décision

gative entre le sentiment de honte et le bien-être individuel (Tangney *et al.*, 2007)¹³. En économie, les analyses récentes de « l'économie du bonheur » révèlent également l'impact des relations sociales ou affectives sur le bien-être.

5. Enfin, la culpabilité (ou la honte) peuvent également avoir des effets comportementaux induits non souhaités. Une étude récente de Hooge, Nelissen, Breugeklans et Zeelenberg (2011) montre ainsi que la culpabilité a des effets bénéfiques pour la personne à qui l'on a fait du tort, mais représente un désavantage pour les autres personnes de notre environnement social. La réparation du tort (en faveur de la victime) se fait au détriment d'autres personnes (des tiers) : un individu se sentant coupable de ne pas s'être suffisamment occupé d'un de ses proches, peut décider de lui consacrer plus de temps pour se rattraper. Cependant, selon les conclusions de Hooge *et al.* (2011), ce temps sera sélectionné sur du temps réservé à un tiers, et non sur du temps réservé à lui-même.

En dépit de l'impact positif des émotions morales à valence négative dans l'émergence des comportements prosociaux, leurs limites augmentent l'intérêt que l'on peut porter aux émotions morales à valence positive, qui ont été pourtant largement moins étudiées jusqu'à présent.

1.3.2. La promotion morale affective des normes sociales ou morales

1.3.2.1. Le rôle de la fierté dans l'accomplissement au travail

La fierté est reliée de façon étroite avec la notion d'accomplissement de soi (Lea et Webley, 1997). L'émotion de fierté a notamment un intérêt, du point de vue de l'économiste et du gestionnaire, dans le cadre de la relation au travail, de la négociation ou lors de l'acceptation des ressources monétaires provenant de dons ou de la charité.

La fierté conduit à des choix qui peuvent sembler a priori irrationnels, lorsque l'individu cherche à effectuer son travail aussi bien que possible (pour pouvoir en être fier) sans pour autant en demander une contrepartie financière, ou lorsqu'il refuse de « s'abaisser » à profiter de dons financiers (familles, œuvres de charité, etc.) qui lui sont pourtant profitables, ou encore lorsqu'il refuse des offres « offensantes » qu'il serait pour autant rationnel, du point de vue de l'homo oeconomicus, d'accepter. Fehr et Schmidt (2005) expliquent, par ailleurs, que la fierté pourrait, au même titre que la colère ou le dégoût influencer le choix de refuser une offre trop faible dans le jeu de l'ultimatum.

Toutefois, les auteurs indiquent que ces décisions, et donc le fait d'être sous l'influence de l'émotion de fierté, peuvent avoir un intérêt pour les individus à long terme. La fierté est notamment liée à l'accomplissement de l'individu dans le travail : elle est une source de motivation intrinsèque (Bénabou et Tirole, 2006), elle renforce les bonnes décisions passées (l'effort au travail par exemple) et est associée à une forte capacité d'autonomie au sein des organisations, ce qui correspond à une stratégie de décision fiable et efficace

13. La tendance générale à ressentir de la honte serait par ailleurs corrélée avec l'occurrence de certaines maladies (dépressions, problèmes liés au poids) ou de certains comportements déviants (abus de drogues, pratiques sexuelles non protégées, arrestations...).

dans un contexte d'incertitude. Dans le cadre d'une négociation, cette émotion établit une réputation positive qui favorise de meilleurs résultats dans le futur. Leur représentation de la fierté est donc celle d'un mécanisme qui aiderait à faire de bonnes décisions à long terme.

Leur travail met en lumière le lien qui existe entre le sentiment de fierté et l'estime de soi. Le choix de ne pas accepter la charité par fierté, ou le choix de refuser une offre offensante en sont des exemples. Cependant, cet aspect de la fierté apparaît plus comme quelque chose de négatif, en lien avec la colère ou avec des mécanismes de défense de l'estime de soi. L'aspect de la fierté qui nous intéresse ici, se trouve lorsque l'on rattache le sentiment de fierté à un plaisir que l'on ressent lorsqu'on réalise soi-même un acte socialement ou moralement valorisé. Dans ce cas, on se place davantage dans un processus de renforcement de l'estime de soi, que dans un processus de défense de l'estime de soi.

Le sentiment de fierté, ou son anticipation, peut donc, à travers la recherche d'un renforcement de l'estime de soi, participer à l'adoption de comportements socialement ou moralement pertinents. Andreoni et Petrie (2004) proposent un jeu du bien public, sans procédure d'anonymat, les joueurs savent alors avec qui ils jouent et savent ce que jouent les autres. Ils expliquent que les forts niveaux de coopération obtenus pourraient être motivés à travers l'influence des émotions de fierté et de honte. Les individus agissent de manière coopérative à la fois dans le but d'éviter le sentiment de honte, provenant du fait d'avoir contribué moins que les autres, et dans le but de ressentir de la fierté d'avoir réalisé une contribution (plus) élevée. On voit ici la complémentarité et l'opposition (parfaite) entre les deux émotions auto-conscientes de honte et de fierté. Ces études questionnent cependant les conditions d'émergence de ces émotions. La survenance des émotions de honte et de fierté et leur impact, semblent en effet être soumis à l'identification complète des partenaires dans le jeu (identité et comportement), puisque les niveaux de coopérations obtenus lorsqu'une seule information sur les deux, ou lorsqu'aucune des informations n'est disponible (anonymat complet), sont significativement moins élevés. Parallèlement, dans l'étude de terrain de Schultz (2007), on observe également que les individus qui réalisent davantage d'économies que les autres et qu'on félicite (via une émoticône souriante) ont tendance à maintenir leur comportement prosocial. L'efficacité de la fierté semble à nouveau être conditionnée à la visibilité de l'acte par autrui.

1.3.2.2. La chaîne vertueuse de la gratitude

A la différence d'un simple sentiment de joie, la gratitude renforce les comportements coopératifs par opposition aux comportements égoïstes lors d'interactions économiques (Barlett et DeSteno, 2006 ; DeSteno, Baumann, Barlett, Dickens et Williams, 2010). DeSteno *et al.* (2010) présentent un jeu similaire au dilemme du prisonnier (se jouant également à deux). Les sujets ayant subi l'induction du sentiment de gratitude – un partenaire lors d'une première tâche, indépendante du jeu, leur prodiguant de l'aide, dans le but de leur faire ressentir de la gratitude – sont plus coopératifs dans le jeu que ceux n'ayant subi aucune induction. De plus, il existe deux conditions pour les individus ayant subi l'induction du sentiment de gratitude. Dans une de ces conditions, on indique aux sujets qu'ils jouent avec le partenaire qui les a aidés à la première étape, tandis que dans l'autre condition, on indique aux sujets qu'ils joueront avec une nouvelle personne inconnue, sans lien avec la tâche précédente. Les auteurs observent alors que les individus ayant subi l'induction du sentiment de gratitude adoptent un comportement plus coopératif dans le jeu, qu'ils jouent avec celui qui les a aidés lors de la première tâche ou bien avec une per-

1. Les émotions morales : un système d'outils au service de la décision

sonne totalement étrangère. L'impact de la gratitude ne pouvant donc pas être attribué à la seule réciprocité entre les agents lorsque l'on considère le deuxième cas.

Ces résultats viennent corroborer ceux d'une étude précédente de Barlett et DeSteno (2006), obtenus en étudiant les comportements d'aide des sujets dans des situations telles que de prendre le temps de répondre à un questionnaire, mais avec la particularité de se placer dans un contexte économique de dilemme social. Les résultats de DeSteno *et al.* (2010) témoignent d'un rôle important de l'émotion de gratitude dans ce contexte spécifique. Ils montrent en effet que cette émotion peut inciter les agents à entreprendre des comportements de réciprocité positive à l'encontre de ceux qui les ont aidés, ce qui constitue un mécanisme de récompense vis-à-vis de ceux-ci, pouvant permettre de promouvoir et de renforcer les comportements prosociaux des bienfaiteurs initiaux. Une étude de Grant et Gino (2010) montre, par ailleurs, qu'une expression de gratitude de la part de celui à qui l'on a donné de l'aide, favorise l'émergence de comportements prosociaux. D'après leurs résultats, cet effet passe par la médiation de la valeur sociale que les individus considèrent qu'on leur attribue. Ainsi, l'expression de gratitude, qu'elle soit un simple remerciement, ou un comportement de réciprocité positive, pousse les individus à se sentir socialement valorisés, éventuellement à travers l'émergence du sentiment de fierté, ce qui entraîne chez eux, le désir de continuer à adopter des comportements prosociaux suscitant la gratitude chez les autres.

L'effet bénéfique de la gratitude dépasse cependant celui de comportements purement réciproques. Dans l'étude de De Steno *et al.* (2010), l'effet fonctionne en effet aussi bien lorsque le sujet joue avec la personne (le complice) qui l'a aidé précédemment que lorsqu'il s'agit d'une autre personne inconnue et sans rapport avec l'aide obtenue par le sujet. Il semble donc que la gratitude puisse provoquer un effet d'entraînement positif en incitant les individus à se comporter de manière prosociale avec ceux qui sont en lien ou non avec la situation d'émergence de cette émotion. La gratitude a donc un effet direct de réciprocité positive mais aussi un effet indirect tourné vers une personne tierce qui n'est pas responsable du sentiment de gratitude. Ce comportement de réciprocité indirecte révèle la capacité de la gratitude à engendrer une « chaîne » de comportements prosociaux (altruisme, coopération, confiance, etc.) qui sont bénéfiques pour les relations interpersonnelles entre les agents au sein de la société. De plus, Grant et Gino (2010) indiquent que l'expression de la gratitude peut également conduire à un renforcement des comportements prosociaux (de ceux à qui elle est exprimée), dirigé à l'encontre de ceux qui expriment la gratitude, mais également vis-à-vis de personnes étrangères au contexte d'émergence de cette émotion. Il semble donc que l'émotion de gratitude provoque des sortes de « spillovers » positifs, en incitant les individus à se comporter de manière prosociale, avec les autres, de manière générale et pas seulement avec ceux qui sont en lien avec la situation d'émergence de cette émotion.

Une étude d'Algoe et Haidt (2008) témoigne également de la capacité de l'émotion de gratitude à participer à la création et au renforcement de relations interpersonnelles entre les agents, avec une étude sur le terrain, examinant les relations existant dans certaines sororités (communautés d'étudiantes) Américaines. Au sein de celles-ci, les sœurs les plus anciennes accueillent les nouvelles entrantes avec des cadeaux. Cet échange se fait par groupe de deux étudiantes, chaque ancienne étant chargée de s'occuper d'une nouvelle. Leurs résultats montrent que le niveau de gratitude résultant de ces comportements a

un impact significatif sur la qualité de la relation d'un couple de sœurs (par exemple, le nombre d'heures passées ensemble dans la semaine). Cet aspect de la gratitude lui confère un rôle prépondérant dans la constitution des réseaux et dans les situations de négociation.

D'après Haidt (2003), enfin, l'émotion d'admiration peut susciter des comportements similaires à ceux de la gratitude, avec un aspect plus général, moins en lien avec une relation bilatérale. Il se pourrait que, similairement au cas de la colère et de l'indignation, la gratitude émerge lorsque l'on est directement affecté par le geste moralement adapté d'un individu, et que l'admiration survienne quant à elle lorsqu'on est simplement témoin de ce type d'actions. Une étude de Schnall, Roper et Fessler (2010) montre que les individus ayant subi l'induction du sentiment d'admiration étaient, comparativement à ceux dans un état émotionnel neutre, plus susceptibles soit de participer volontairement à une étude non payée ou à aider l'expérimentateur à réaliser une tâche difficile, soit d'adopter un comportement prosocial dans un jeu de dilemme social. Néanmoins, il n'existe pas d'étude en économie examinant le rôle de cette émotion sur le comportement des agents.

1.4. Conclusion

Notre revue de la littérature a permis de mettre en perspective une abondante littérature intégrant, dans l'analyse du processus de décision économique, le rôle des émotions morales négatives ou positives. Comme l'anticipait Frank (1988), les émotions morales à valence négative ont la capacité d'influencer les comportements stratégiques des individus en les plaçant dans une optique de long terme. La culpabilité ou la honte s'opposent à la recherche de l'intérêt personnel immédiat en laissant augurer les conséquences émotionnelles négatives de comportements opportunistes. Ces émotions morales participent ainsi de la régulation à long terme des comportements sociaux comme la coopération, la confiance ou le sens de l'équité. Notre revue souligne cependant les limites de ces émotions morales à valence négative en termes de bien-être et parfois d'efficacité. L'apport des émotions morales positives, comme la gratitude ou l'empathie, en est ainsi renforcé. Les émotions positives, à travers leur rôle dans la promotion et la diffusion des comportements prosociaux entre les individus, méritent davantage d'attention de la part des économistes et des gestionnaires.

2. La Gratitude et le Côté Radieux des Emotions Morales

2.1. Introduction

En dépit d'un intérêt précoce par des philosophes comme Cicéron (1851) ou Adam Smith (1759), la gratitude a été plutôt négligée par la psychologie moderne, comme par l'économie, jusqu'au début du nouveau millénaire (McCullough, Kilpatrick, Emmons et Larson, 2001). La gratitude est définie comme l'émotion que l'on éprouve lorsque l'on bénéficie d'une action coûteuse entreprise par un autre individu, et qui est réalisée intentionnellement et volontairement, sans aucune compensation financière immédiate pour le bienfaiteur (McCullough *et al.*, 2001).

La gratitude, qui fut considérée comme la principale vertu chez Cicéron (1851), a été pendant longtemps confinée parmi d'autres émotions à valence positive également peu étudiées (comme la joie ou l'admiration). Ce phénomène provient du fait que les chercheurs ont plutôt mis l'accent sur diverses émotions à valence négative comme la culpabilité, la colère ou la honte, car celles-ci sont supposées avoir une plus grande influence sur les comportements humains (Baumeister, Bratslavsky, Finkenauer et Vohs, 2001), sans doute parce que cette influence peut, en quelques sortes, être perçue de façon plus intense. D'une manière générale, les réactions à des événements négatifs sont censées être plus fortes que les réactions à des expériences plus positives parce que les premières sont considérées comme promulguant une meilleure capacité d'adaptation au niveau évolutif. Ce type de raisonnement s'étend également à l'émergence des émotions, qui sont des réactions humaines naturelles à la survenance d'événements négatifs et positifs. En effet, la promotion de la capacité d'adaptation d'un individu à son environnement peut sembler plus en lien avec le fait de savoir faire face aux menaces éventuelles, plutôt qu'avec le fait d'embrasser de nouvelles opportunités. Parallèlement, les difficultés auxquelles les individus doivent répondre nécessitent une intervention afin de parvenir à la résolution du problème en question, tandis que, lors de la survenance d'événements positifs, aucune intervention n'est nécessaire, ce qui attire logiquement moins d'attention.

Cette domination, en termes de niveau d'attention, de la part des événements, expériences et réactions négatives, par rapport à celles plus positives, se retrouve dans diverses facettes de la psychologie humaine. Par exemple, elle est présente dans les processus d'apprentissage (la punition pouvant être considérée comme plus marquante que la récompense), la santé, l'attitude envers les risques ou également dans l'analyse des interactions sociales et de l'évolution des relations interpersonnelles au cours du temps. Ainsi, Baumeister *et al.* (2001) ont noté que le succès à long terme d'une relation (entre deux amants) dépend plus de la capacité des partenaires à éviter de faire de mauvaises choses (qui nuisent à l'autre personne) que sur leur capacité à parvenir à faire de bonnes choses.

2. La Gratitude et le Côté Radieux des Émotions Morales

Néanmoins, ces auteurs reconnaissent également la puissance du bien : « Le bon peut triompher du mauvais par la force du nombre. Pour maximiser le pouvoir des bonnes choses, leur nombre doit être augmenté. Ceci est possible en créant plus de bonnes choses. Par exemple, dans une relation amoureuse, chaque partenaire peut faire l'effort d'être systématiquement gentil envers l'autre. De tels petits actes de gentillesse sont importants afin de combattre les mauvaises choses qui surviennent généralement. (...) De même, les individus peuvent faire l'effort de reconnaître et d'apprécier les bonnes choses qui leur arrivent - en célébrant chaque petit succès, en étant reconnaissant d'être en bonne santé et en ayant de la gratitude envers ceux qui les soutiennent » (Baumeister *et al.*, 2001, p. 361, notre traduction). En outre, la littérature en psychologie positive s'applique également à défendre le rôle adaptatif des émotions positives comme la gratitude.

En particulier, la théorie « *broaden-and-build* » de Fredrickson's (2001, 2004) suggère que les émotions positives comme la joie, l'intérêt, le contentement, l'amour, la fierté ou l'admiration fonctionnent ensemble à court terme dans le but d'élargir l'esprit et la pensée des individus, leur permettant de s'ouvrir à de nouvelles ressources cognitives, sociales, psychologiques, voire physiques, à plus long terme (Vacharkulksemsuk, Sekerka et Fredrickson, 2011). Plus précisément, cette théorie affirme que les émotions positives suscitent des comportements qui sont moins spécifiquement orientés vers le stimulus que les émotions négatives, lesquelles sont souvent étroitement liées à la situation stimulante et mènent par la suite à un panel de tendances à l'action plus restreint. Par exemple, la joie génère chez l'individu l'envie d'être créatif ; l'intérêt stimule l'envie d'explorer et de réaliser des expériences ; tandis que la fierté pousse à envisager de nouvelles réalisations à venir, autant qu'à partager les précédentes réussites avec son entourage (Fredrickson, 2004).

Ainsi, les comportements et les tendances à l'action provoquées par les émotions positives comportent des avantages adaptatifs à long terme grâce à la possibilité de tirer profit des ressources personnelles durables qui sont développées, au cours du temps, sur leur base. Ces ressources personnelles sont aussi bien d'ordre physique et intellectuel que d'ordre psychologique ou social. Par exemple, l'exploration poussée par l'intérêt crée de la connaissance et de la complexité intellectuelle, tandis que la fierté nourrit l'estime de soi et la motivation à la réussite (Fredrickson, 2004).

La gratitude, en tant qu'émotion positive, a également été intégrée par Fredrickson dans la théorie « *broaden-and-build* » et possède, en conséquence, une influence certaine sur la valeur sélective (*fitness*) des êtres humains face à leur environnement, dans la mesure où cette émotion aide à construire et à renforcer les relations sociales et les amitiés (Emmons et Shelton, 2002) qui constituent des ressources décisives pour les agents. Cet enseignement fait directement écho au rôle de la gratitude comme décrit par Adam Smith dans « *La théorie des sentiments moraux* » (1759). Dans ce travail séminal, il a dépeint la gratitude comme essentielle au fonctionnement de la vie sociale en raison de sa capacité à promouvoir une plus grande stabilité des relations au sein de la société.

2.2. La gratitude et les relations interpersonnelles

Le rôle crucial de la gratitude dans les relations interpersonnelles est d'un intérêt primordial pour les économistes. Ceci est particulièrement vrai pour les économistes dont l'analyse est centrée sur la coopération et ses déterminants, en particulier dans le contexte des dilemmes sociaux. Les dilemmes sociaux sont des situations d'interaction dans lesquelles les agents sont tiraillés entre leur intérêt personnel (immédiat) et l'intérêt collectif (généralement envisagé à plus long terme) (Dawes, 1980 ; van Lange et Joireman, 2008). Aux côtés d'autres émotions morales comme la culpabilité, la colère, la honte ou la fierté, la gratitude joue un rôle clé dans la façon dont les interactions sociales opèrent dans le contexte spécifique des dilemmes sociaux, principalement par le biais de sa capacité à établir et à consolider des liens étroits entre les individus. De nombreux travaux empiriques et théoriques, à la fois dans les champs de la psychologie et de l'économie, ont mis en lumière plusieurs caractéristiques de la gratitude à travers lesquelles cette émotion est en mesure de remplir une fonction dans le bon déroulement des relations interpersonnelles.

2.2.1. La réciprocité positive

Tout d'abord, on peut évidemment évoquer la principale tendance à l'action de la gratitude, soit l'incitation à la mise en place d'une réciprocité positive entre les agents. En effet, la gratitude a été désignée par de nombreux économistes, sociologues et psychologues sociaux, comme la principale source de réciprocité positive (voir par exemple Simmel, 1950). S'agissant de la mise en place de la réciprocité positive, il convient de discriminer entre deux de ses formes habituelles. Premièrement, nous trouvons la réciprocité positive directe, qui conduit le bénéficiaire d'une bonne action à ressentir le besoin de rendre la faveur à son bienfaiteur, même lorsque cela est coûteux. Deuxièmement, on parle de réciprocité positive indirecte, lorsque le besoin d'agir généreusement, en retour d'une action prosociale, se matérialise finalement sous la forme d'un geste bienveillant à l'encontre d'un tiers, non impliqué dans la situation initiale conduisant à l'émergence de la gratitude¹. De nombreuses études empiriques ont systématiquement indiqué que la réciprocité positive (qu'il s'agisse de sa forme directe ou indirecte) est significativement et positivement influencée par le sentiment de gratitude, à la différence d'autres émotions ou humeurs (surtout positives) (Tsang, 2006b ; Bartlett et DeSteno, 2006 ; DeSteno, Baumann, Bartlett, Dickens et Williams, 2010).²

En particulier, Tsang (2006b) souligne la distinction entre la gratitude et le senti-

1. Pour être complet, il convient de noter qu'il est possible d'observer la réciprocité indirecte sous deux formes différentes. En effet, comme cela est expliqué dans Nowak et Roch (2007), on trouve d'une part la réciprocité indirecte « en aval » (notre traduction), qui exprime le fait qu'une personne qui a aidé, ou qui a agi d'une manière prosociale, dans le passé, a une plus grande chance de recevoir de l'aide, ou de bénéficier d'un acte prosocial à son tour (la plupart du temps en raison de considérations de réputation). D'autre part, il existe également la réciprocité indirecte « en amont » (notre traduction), qui implique qu'une personne qui vient de recevoir l'aide, ou qui vient de bénéficier d'un acte prosocial, a envie d'aider ou d'agir d'une manière prosociale à son tour, envers quelqu'un d'autre que le bienfaiteur initial. Ici, nous nous concentrons sur la deuxième forme de réciprocité indirecte, puisque celle-ci est plus en lien avec l'influence du sentiment de gratitude. Ainsi, dans le reste de la thèse, lorsque l'expression réciprocité indirecte est mentionnée, cela se réfère à la réciprocité indirecte « en amont ».

2. Dans le cas de la réciprocité indirecte, le rôle de la gratitude s'étend même à des êtres non humains. En effet, certains travaux ont pu montrer que les gens qui rendent plus fréquemment grâce à Dieu sont plus susceptibles d'aider les autres (par exemple, Krause, Evans, Powers et Hayward, 2012).

2. La Gratitude et le Côté Radieux des Emotions Morales

ment d'endettement, qui est une autre réaction possible suite à l'obtention d'une faveur. L'endettement est défini comme une sensation désagréable conduisant les individus à se concentrer sur l'obligation sociale qu'ils ont de rembourser leur bienfaiteur. Ce sentiment provoque alors généralement la peur d'être incapable de mener à bien le souhait de remboursement de l'agent, et il est souvent accompagné d'une anxiété chez celui-ci, due à la recherche de moyens permettant de rembourser le bienfaiteur (McCullough *et al.*, 2001). Tsang (2006b) soutient que ces deux sentiments (gratitude et endettement) peuvent aider à l'évaluation de la valeur d'un présent. De même, ces deux réactions émotionnelles peuvent fonctionner comme un rappel de la présence de la norme sociale de réciprocité. Néanmoins, selon lui, la gratitude et le sentiment de dette fournissent une information différente à l'individu, s'agissant du fait de valoriser, ou non, le fait d'être sujet à la norme de réciprocité, dans une situation donnée, et que cela pourrait affecter la capacité de l'individu à se conformer à cette norme. Plus précisément, le fait qu'un individu ressente de la gratitude, une émotion à valence positive, pourrait indiquer que celui-ci a ressenti de manière positive le fait de recevoir un bénéfice et qu'il développe des sentiments positifs à l'égard de son bienfaiteur, ce qui augure, chez lui, une plus grande facilité à faire preuve de réciprocité positive, par la suite.

De façon complémentaire, d'autres travaux ont relié la distinction entre la gratitude et le sentiment d'endettement aux intentions des bienfaiteurs. En effet, il semble que la gratitude émerge à la suite d'un acte prosocial, lorsque cette action est perçue comme le signe d'une considération positive de la part du bienfaiteur, alors que l'endettement intervient lorsque les intentions de celui-ci sont considérées comme manipulatrices ou trompeuses (Greenberg et Wescott 1983). En outre, Tsang (2006a) a montré, à l'aide d'une série d'études expérimentales, que l'émotion de gratitude est plus sensible aux intentions d'un bienfaiteur que ne l'est l'endettement. Plus précisément, l'auteur explique que les personnes ont ressenti plus de gratitude en réponse à une faveur qui leur accordée avec des intentions bienveillantes. En revanche, le sentiment de dette, développé par les individus, n'est pas lié à la nature des intentions de la personne qui les a aidé. Selon Tsang (2006a), le fait que le sentiment de dette ne soit pas affecté par les intentions du bienfaiteur pourrait provenir de l'influence de la norme de réciprocité, puisque celle-ci s'applique toujours, peu importe si l'aide du bienfaiteur a été prodiguée avec des intentions bienveillantes ou non.

En outre, Watkins, Scheer, Ovnicek, et Kolts (2006) ont également mis l'accent sur les attentes de réciprocité du bienfaiteur. En utilisant des scénarios, ils ont constaté que l'expression d'attentes, de la part d'un bienfaiteur, vis-à-vis de la mise en place d'une réciprocité positive, après un acte prosocial, est négativement corrélée avec le niveau de sentiment de gratitude déclaré par les participants, tandis que la relation inverse a pu être mise en évidence concernant le sentiment d'endettement. Par ailleurs, les participants sont plus enclins à rapporter la volonté de retourner une faveur lorsque le bienfaiteur n'a pas exprimé d'attentes fortes vis-à-vis de la mise en place d'une réciprocité positive.

Considérées dans leur ensemble, ces résultats semblent indiquer que, pour que la gratitude puisse émerger, les intentions du bienfaiteur doivent être bienveillantes et, en particulier, les attentes de réciprocité (susceptibles de communiquer des intentions manipulatrices) ne doivent pas être rendues saillantes. Dans ces conditions, lorsque la gratitude l'emporte sur le sentiment d'endettement, les agents semblent apprécier le fait d'être soumis à la norme de réciprocité et, par conséquent, sont plus enclins à se conformer à cette

norme en retournant une faveur.

En plus d'avoir une influence sur l'établissement d'une réciprocité positive entre les agents, motivant l'adoption de comportements prosociaux chez les bénéficiaires d'un acte bienveillant, la gratitude a également été identifiée comme permettant un renforcement de ces comportements prosociaux chez les bienfaiteurs (McCullough *et al.*, 2001). Ainsi, Grant et Gino (2010) ont notamment montré que les bienfaiteurs, lorsqu'ils sont remerciés (que ce soit par un geste réciproque ou par la reconnaissance que leur geste initial a été apprécié), ont tendance à renouveler leurs comportements prosociaux à l'avenir, que ce soit vis-à-vis du même bénéficiaire ou d'un tiers.

En somme, la gratitude, dans le contexte des comportements réciproques, peut être une force motrice en vue de l'adoption de gestes prosociaux à travers de multiples canaux. D'une part, l'émergence de la gratitude peut provoquer, chez le destinataire d'un acte prosocial, l'envie de « rendre la pareille », que ce soit directement (à destination de son bienfaiteur) ou indirectement (à un tiers). D'autre part, l'expression de la gratitude tend à valider le comportement prosocial d'un bienfaiteur, qui est, par conséquent, renforcé dans sa volonté d'agir avec de manière bienveillante, que ce soit envers le bénéficiaire initial ou envers un tiers.

L'aspect indirect qui caractérise l'influence de la gratitude semble indiquer que cette émotion va au-delà des considérations de réciprocité et est susceptible de créer des cercles vertueux et des effets d'entraînement entre les individus, les aidant à développer des relations courtoises et permettant à la coopération d'être cimentée au sein des groupes d'agents, à travers une dynamique positive. Un tel résultat viendrait renforcer l'idée, notamment envisagée par Adam Smith (1759), selon laquelle la gratitude joue un rôle crucial au niveau relationnel. Dans la section suivante, nous examinons plus en détail l'impact de la gratitude sur les relations sociales.

2.2.2. L'influence de la gratitude sur la qualité des relations entre agents

2.2.2.1. Proximité et engagement

Comme l'affirment Algoe, Gable et Maisel (2010, p. 217, notre traduction), « une caractéristique déterminante des relations intimes entre adultes est que chaque membre effectue des actions qui bénéficient à l'autre ». Cette citation nous permet d'illustrer l'idée que les agents peuvent devenir plus proches à travers un enchaînement réciproque de gestes bienveillants, assujettis à l'émergence de la gratitude. Ainsi, cette émotion remplit une fonction évolutive chez les êtres humains qui a été introduite dans la théorie « Find-remind-and-bind » proposée par Algoe, Haidt et Gable (2008) et entièrement développée par Algoe (2012). En effet, la gratitude joue un rôle clé dans la formation du réseau des individus, qui sont par essence des êtres sociaux. Par exemple, la gratitude est en mesure d'orienter les individus vers des partenaires potentiellement plus rares et bénéfiques que les autres, dans la mesure où cette émotion donne aux agents l'opportunité de solidifier la connexion qu'ils ont avec ceux qui leur apportent du soutien et qui enrichissent leur vie (Algoe, 2012). De cette façon, dans le contexte des relations amoureuses, Algoe *et al.* (2010, p. 221, notre traduction) expliquent que : « Les moments de gratitude peuvent agir comme des « piqûres de rappel » au cours d'une relation : la gratitude aide un individu à

2. La Gratitude et le Côté Radieux des Emotions Morales

se rappeler de ses sentiments à l'égard de son partenaire et inspire une sensibilité mutuelle qui sert à accroître le lien existant au sein du couple ».

En d'autres termes, l'émergence de la gratitude permet aux agents d'identifier et de se rappeler des personnes qui ont une influence positive dans leur vie et cela les motive également à s'investir dans ces relations (par des actes prosociaux) afin de se rapprocher et/ou, au moins, de parvenir à rester proches de ces partenaires particulièrement précieux, à mesure que ceux-ci deviennent parties intégrantes de leurs ressources sociales et psychologiques, au fil du temps.

2.2.2.1.1. Au sein des relations dyadiques

Le travail de Goeree, McConnell, Mitchell, Tromp et Yariv (2010), qui combine des données sur les réseaux d'amitié et les caractéristiques individuelles de sujets avec l'observation expérimentale de leurs comportements dans le jeu du dictateur, se rapporte à ces phénomènes. Ils ont constaté que le niveau des offres des dictateurs s'explique principalement par la distance sociale. D'ailleurs, ils ont détecté que les sujets (élèves états-unien dans des classes de primaire et de secondaire) affichent des comportements d'homophilie, à savoir qu'ils ont tendance à vouloir se connecter avec des personnes qui leur sont semblables.

Les comportements d'homophilie observés dans cette étude indiquent que lorsque les agents cherchent à identifier de potentiels partenaires de valeur, comme extensions de leur réseau, ils ont tendance à favoriser les individus avec lesquels ils partagent des similitudes. L'étude a porté principalement sur des caractéristiques physiques telles que la race ou la taille³, bien que le processus d'identification des partenaires souhaités (potentiellement utiles) puisse s'étendre également à d'autres types de caractéristiques tels que des valeurs similaires (comme l'orientation politique et éventuellement des normes morales partagées) (Huston et Levinger, 1978 ; McPherson, Smith-Lovin et Cook, 2001), des tendances de comportements similaires (y compris pour les comportements prosociaux) (McPherson *et al.*, 2001), ainsi que, sans doute, des dispositions similaires pour ce qui est de ressentir et d'exprimer la gratitude.

En outre, les données de cette étude ont montré que les comportements de don suivent une loi inverse par rapport à la distance sociale, ce qui vient corroborer la théorie « Find-remind-and-bind » vue précédemment. En effet, cela illustre le désir des individus de donner la priorité à l'amélioration des relations qu'ils ont déjà définies comme précieuses (ou au moins souhaitées), lorsque la possibilité d'être généreux leur est donnée. En conséquence, nous interprétons les résultats obtenus par Goeree *et al.* (2010) comme la mise en évidence d'une étape isolée au sein d'une chaîne vertueuse de gestes bienveillants, déjà initiée avant l'étude, étant donné que des relations étroites avaient déjà été établies entre certains sujets. Cette façon de procéder est conforme avec la tendance de comportement que Bartlett, Condon, Cruz, Baumann et DeSteno (2012) ont attribué à la gratitude, vis-à-vis de la promotion des relations en cours, comme des relations qui sont nouvellement formées, soit le fait de choisir de s'affilier avec un partenaire et d'agir de manière solidaire avec lui même lorsqu'une telle action est coûteuse. Pris dans leur ensemble, ces éléments

3. Pour une revue sur l'impact de l'homophilie sur la formation du réseau, voir McPherson, Smith-Lovin et Cook (2001). Les auteurs examinent différents types d'homophilie, de la race (origine ethnique) à l'âge, à la religion, à l'éducation, à la profession et au sexe.

suggèrent que la gratitude favorise la proximité entre les agents dans une relation, ainsi que l'engagement des partenaires l'un envers l'autre.

Par ailleurs, il a également été démontré que l'émotion de gratitude, lorsqu'elle est *exprimée* à un partenaire, intervient au niveau de sa perception de la force collective (de l'anglais « *communal strength* », notre traduction) de la relation (Lambert, Clark, Durtschi, Fincham et Graham, 2010). La force collective peut être définie comme le degré de responsabilité ressenti vis-à-vis du bien-être d'un partenaire (Mills, Clark, Ford et Johnson, 2004) et peut être considérée comme une dimension caractérisant la proximité d'une relation entre deux agents. Schématiquement, plus la force collective est importante au sein d'une relation, plus les partenaires qui la constituent sont considérés comme proches. La force collective est également liée à la définition des relations intimes (ou commune, de l'anglais « *communal relationships* », notre traduction), que l'on différencie des relations dites d'échange. La catégorie des relations intimes englobe l'ensemble des relations pour lesquelles un individu a un sentiment de responsabilité à l'égard de son partenaire, notamment quant à la recherche de moyens pour subvenir à ses besoins, sachant que les bienfaits prodigués au partenaire sont délivrés en fonction de ces besoins.

A l'inverse, les relations dites d'échange impliquent des relations dans lesquelles les bienfaits (que l'on peut qualifier de prestations) sont fournis en échange d'autres bienfaits, et ne sont pas subordonnés à la nécessité du destinataire (Clark et Mills, 2011). Il est intéressant de noter que les études menées par Lambert, Clark, Durtschi, Fincham et Graham (2010) ont montré que l'expression de la gratitude améliore la perception de la force collective dans les relations dyadiques, notamment au fil du temps. Ainsi, l'expression de la gratitude semble jouer un rôle clé dans la dynamique évolutive des relations et, plus précisément, dans la façon dont les agents se lient et s'engagent auprès de partenaires précieux au cours du temps. Ainsi, ce résultat permet de mieux comprendre ce qui rend la gratitude et son expression précieuses dans la création et le renforcement des liens sociaux.

2.2.2.1.2. Au sein du groupe

Au delà de son effet au niveau des relations dyadiques, Algoe *et al.* (2008) ont constaté que l'influence de la gratitude peut aussi être perçue au niveau du groupe. On rappelle que les résultats obtenus par les auteurs montrent que chez de nouvelles venues dans des sororités américaines, la gratitude issue du fait d'être accueillies avec des cadeaux de bienvenue, offerts par une sœur aînée (chaque nouvelle arrivante étant appariée avec une sœur aînée), a été associée à un plus grand sentiment d'intégration au groupe. Ainsi, l'influence de la gratitude sur les relations, bien que résultant le plus souvent d'interactions au niveau dyadique, peut également s'opérer au niveau du groupe. Cet aspect fait écho aux mécanismes de réciprocité positive indirecte, également en lien avec l'émotion de gratitude, que nous avons évoqués ci-dessus et cela renforce l'idée que cette émotion peut être la source d'effets d'entraînement profitants aux groupes d'individus et à la société dans son ensemble.

En outre, les émotions positives, comme la gratitude, changent la façon dont les gens se voient, élargissant ainsi le champ de leur perception d'eux-même, afin d'y inclure leurs proches (Vaughn et Fredrickson, 2006). Sur cette question, Vacharkulksemsuk *et al.* (2011, p. 107, notre traduction) disent : « Alors que la ligne permettant de distinguer entre soi-même et les autres devient trouble et plus difficile à délimiter, les individus peuvent

commencer à adopter intérieurement les caractéristiques des autres, envisageant l'autre personne comme acceptable ou comme faisant part de leur propre répertoire. Dans la mesure où les individus voient leurs collègues ou leur organisation comme faisant partie d'eux-même (...) l'allocation de ressource peut être perçue comme une responsabilité commune plutôt qu'individuelle ». A cet égard, la gratitude, en tant qu'émotion positive, contribue à renforcer le partenariat entre les agents, ce qui est essentiel pour la promotion de la coopération. Par ailleurs, le phénomène décrit par Vacharkulksemsuk *et al.* (2011) semble indiquer qu'un climat émotionnel positif permet aux individus d'adopter des relations intimes les uns avec les autres, dont le maintien et le renforcement, au cours du temps, sont favorisés grâce à l'émergence et à l'expression de gratitude.

En bref, la gratitude participe à la fois à la création d'un sentiment d'appartenance à un groupe (« *Je fais partie de ce groupe, je sens que les individus de ce groupe se préoccupent de moi.* »), ainsi qu'à la promotion de la prise de conscience que d'autres individus sont également inclus dans le groupe et qu'il convient de leur donner de la considération (« *Ces individus font partie du groupe, je me préoccupe d'eux également.* »). Ceci lie clairement la gratitude au développement des relations entre les agents au niveau du groupe.

2.2.2.2. Confiance et satisfaction

Une autre dimension primaire dans l'analyse des relations sociales et de leur évolution dynamique est le niveau de confiance établie entre les différentes parties. La confiance a été définie par Dunn et Schweitzer (2005, p. 736, notre traduction) comme « la volonté d'accepter d'être en position de vulnérabilité, basée sur des attentes positives concernant le comportement des autres ». En ce sens, nous soutenons que, pour être prêt à se comporter de manière prosociale, par exemple en fournissant de l'aide à un autre individu, en s'investissant dans un projet de collaboration ou encore en contribuant à un bien public, il est nécessaire d'avoir suffisamment confiance dans les autres. En effet, le fait de réaliser de tels actes bienveillants implique que l'on se mette dans une position de vulnérabilité où l'on risque d'être exploité par les autres, comme, par exemple, lorsque l'on est confronté à des comportements de « passagers clandestins ». Autrement dit, faire confiance à un partenaire semble être une condition nécessaire à la bonne mise en œuvre d'actes de bienfaisance, qui sont nécessaires au développement favorable des relations. Dans la littérature, il est suggéré que le niveau de confiance entre les partenaires est soumis à l'influence de la gratitude. Ainsi, Komter (2004, p. 211, notre traduction) indique que « sans les liens créés par la gratitude, il n'y aurait pas de confiance mutuelle, pas de bases morales sur lesquelles agir, et pas de motif justifiant de maintenir l'engagement au sein d'une communauté ».

Une explication possible à l'influence de la gratitude sur la confiance est que les sentiments peuvent affecter les jugements et, en particulier, il apparaît que l'évaluation que font les individus au sujet de la fiabilité d'une autre personne procède souvent par le biais de l'examen des sentiments qu'ils ont envers cette personne en particulier (Jones et George, 1998). De même, McAllister (1995) suggère que la confiance est enracinée dans les liens affectifs et suppose que les comportements altruistes fournissent un support pour la part de la confiance qui est fondée sur l'affect. Or, la gratitude, comme nous l'avons mentionné ci-dessus, tend à agir en faveur du développement de sentiments bienveillants entre les agents, notamment par la réciprocité des comportements altruistes. Par conséquent, la gratitude participe probablement à l'amélioration de la perception de fiabilité

que les agents ont les uns envers les autres (Palmatier, Jarvis, Bechkoff et Kardes, 2009) et favorise la création de relations de confiance entre les individus.

Plusieurs études empiriques ont apporté des preuves crédibles soutenant une telle proposition. En particulier, Dunn et Schweitzer (2005) ont montré que les personnes chez qui le sentiment de gratitude a été induit, par le souvenir, ont plus tard rapporté des niveaux plus élevés de confiance envers un tiers que ne l'ont fait les participants pour lesquels l'émotion induite était la colère, la culpabilité ou la fierté. De façon intéressante, les auteurs ont également constaté que l'influence de la gratitude sur la confiance dépend de l'histoire partagée par les individus en question, de sorte que cette influence n'est pas significative dans le cas d'un tiers connu, sans doute parce qu'il a déjà été évalué comme (in)digne de confiance, sur la base d'expériences passées. Ceci tend à suggérer que l'influence de la gratitude peut être encore plus cruciale dans les premiers moments de la formation d'une relation, en tant que moyen d'orienter le développement de celle-ci sur un chemin propice, au moins en ce qui concerne le niveau de confiance entre les partenaires.

De plus, dans le contexte de la relation client/vendeur, Palmatier *et al.* (2009) ont obtenu des données, à la fois en laboratoire et sur le terrain, montrant que le sentiment de gratitude des clients (par exemple, après avoir bénéficié d'attentions particulières) affecte positivement leur niveau de confiance envers les vendeurs. En outre, Jackson, Lewandowski, Fleury et Chin (2001) ont constaté que l'induction de la gratitude chez des participants les amènent à attribuer la bonne fortune d'une autre personne à des causes jugées comme stables et contrôlées par la personne chanceuse. Ainsi, selon McCullough, Kimeldorf et Cohen (2008), ces résultats fournissent une preuve que la gratitude conduit les individus à donner du crédit aux autres pour leurs réalisations. En conséquence, nous pensons que la gratitude peut aussi favoriser le niveau de confiance dans la capacité des gens (en plus de leurs intentions) à se comporter de manière appropriée au cours des interactions sociales.

Enfin, Wood, Joseph et Maltby (2008) ont constaté que la gratitude, comme trait de personnalité, est liée à plusieurs facettes de personnalité que l'on retrouve chez les êtres humains et qui sont indicatives de la qualité d'une relation, soit, la confiance, l'altruisme et l'idéalisme⁴. Ceci contribue à véhiculer l'idée d'une association systématique entre la gratitude et la confiance.

Finalement, il est intéressant de noter que plusieurs études ont également lié la gratitude à la qualité des relations de long terme (Bartlett *et al.*, 2012). En particulier, l'expérience de la gratitude est positivement corrélée avec la satisfaction des partenaires au sein de relations dyadiques avec un fort niveau d'engagement mutuel (Algoe *et al.*, 2010).

En somme, que ce soit au niveau dyadique ou au niveau du groupe, la gratitude aide les agents à : (1) trouver des partenaires précieux, (2) se rapprocher de ces partenaires et (3) s'engager avec eux, ce notamment à travers (4) un niveau de confiance accru, pour finalement (5) bénéficier de la satisfaction de maintenir des relations de qualité avec eux.

4. En psychologie, la personnalité des êtres humains est caractérisée par cinq principaux domaines psychologiques : extraversion, ouverture, amabilité, conscience et névrosisme. De plus, chaque domaine est composé de six facettes. Par exemple, la confiance appartient au domaine psychologique de l'amabilité, aux côtés de la droiture, de l'altruisme, du respect, de la modestie et de l'idéalisme.

La gratitude joue ainsi un rôle clé dans la création des relations sociales, ainsi que dans l'amélioration de leur qualité sur le long terme, notamment grâce à sa capacité à initier et à favoriser des cercles vertueux de comportements bienveillants entre les agents, ce qui rend cette émotion particulièrement utile durant les bons moments d'une relation.

Néanmoins, il peut y avoir des moments plus difficiles pendant lesquels les relations peuvent être menacées par la survenance d'événements interpersonnels négatifs comme l'adoption de comportements égoïstes ou de passagers clandestins par un partenaire. En ce qui concerne ces différentes menaces, un autre aspect intéressant de la gratitude est son rôle dans la protection et la conservation des relations entre les agents.

2.2.3. La gratitude comme un moyen de protéger et de préserver les relations

2.2.3.1. Le rôle *ex ante* de la patience

L'une des meilleures façons de lutter contre les événements négatifs interpersonnels et leurs conséquences est d'empêcher jusqu'à leur simple apparition. Or, ces événements, quand ils ne sont pas le fruit du hasard, prennent généralement la forme de choix stratégiques favorisant des récompenses immédiates par rapport à des gratification plus différées. Tel est le cas, par exemple, lorsqu'une personne décide de profiter d'un partenaire, en négligeant le fait que cela pourrait se faire au détriment de l'accès aux avantages de cette relation à l'avenir. Par conséquent, la plupart du temps, afin d'éviter les événements interpersonnels négatifs, il est nécessaire que les agents soient moins impatients.

En effet, la patience est une vertu qui peut aider les agents à mener une vie sociale réussie en favorisant l'acceptation des sacrifices à court terme en échange de gains futurs. En particulier, « Les bénéfices provenant de la coopération et de la confiance, par exemple, requièrent que l'on accepte les coûts immédiats à fournir de l'aide à une autre personne afin de pouvoir profiter, en retour, des gains de long-terme associés à une relation qui perdure et qui est caractérisée par des échanges continus » (DeSteno, Li, Dickens et Lerner, 2014, p. 1263, notre traduction). On pourrait considérer le degré de patience d'un agent comme quelque chose qui est fixe et immuable, quelque chose qui est lié à sa personnalité. Cependant, DeSteno *et al.* (2014) ont constaté que l'émotion de reconnaissance (par opposition à l'état positif plus général du bonheur) réduit l'impatience économique. Ce résultat implique que l'émotion de gratitude, par son influence sur le niveau de la patience des individus, participe à la prévention des comportements qui peuvent menacer la qualité des relations. Ceci est illustratif de la capacité de la gratitude à protéger les relations, avant même que quelque chose de mal ne puisse se produire entre les partenaires.

Néanmoins, qu'en est-il lorsque des événements interpersonnels négatifs se produisent malgré tout ? Est-ce que la gratitude est liée à un processus de guérison des relations endommagées ? La gratitude peut-elle œuvrer contre la mise en place de cercles vicieux entre partenaires contrariés, prenant la forme de comportements malveillants réciproques (à des fins de vengeance), et qui conduisent généralement à la fin d'une relation ? Nous soutenons que la gratitude peut aider à guérir les relations, suite la survenance d'événements interpersonnels négatifs, à travers son influence sur la capacité des agents à pardonner.

2.2.3.2. Le rôle *ex post* du pardon

Le pardon est défini comme une réponse à des événements interpersonnels négatifs. Les événements négatifs motivent généralement à fuir une relation ou à chercher à se venger, soit à entreprendre des comportements qui menacent la qualité d'une relation, sinon, sa simple survie. McCullough (2000) présente une conceptualisation motivationnelle qui considère le pardon comme la possibilité de passer de motivations destructrices pour les relations (évitement, vengeance) à des motivations plus (re)constructrices des relations, en suscitant des comportements tels que les tentatives de réconciliation, et en inhibant l'apparition d'agressions ou de représailles. En ce sens, le pardon peut être considéré comme un changement psychologique prosocial.

McCullough (2000) identifie plusieurs déterminants du pardon. Parmi ceux-ci, on trouve le processus émotionnel de l'empathie, qui est lié à la capacité de pouvoir s'imaginer à la place de quelqu'un d'autre ; la tendance à ruminer et à essayer de supprimer la rumination, qui influence négativement la capacité des agents à pardonner ; des facteurs conjoncturels tels que les excuses ; et enfin, certaines qualités relationnelles telles que la proximité entre les partenaires, l'engagement qu'ils ont l'un envers l'autre, ainsi que la satisfaction avec laquelle ces individus entretiennent leur relation.

Tout d'abord, l'empathie permet aux agents de comprendre la perspective cognitive d'un offenseur, ce qui est fortement corrélée avec des mesures globales du pardon (McCullough, Worthington Jr et Rachal, 1997). Fait intéressant, l'émergence de la gratitude est également liée aux dispositions empathiques des individus, puisque la compréhension des intentions bienveillantes d'un bienfaiteur est nécessaire pour éveiller le sentiment de gratitude (Lazarus et Lazarus, 1994).

Toutefois, l'existence d'un lien entre la proximité, l'engagement et la satisfaction au sein d'une relation et la capacité des partenaires à se pardonner les uns les autres, est encore plus intéressant. Il est évident que des relations de qualité sont très précieuses pour les êtres humains en tant qu'individus sociaux. Ainsi, il semble raisonnable de supposer que plus une relation entre deux agents est caractérisée par la proximité, l'engagement et la satisfaction, plus ces personnes seront affectées par sa destruction. En conséquence, nos relations les plus importantes doivent être protégées contre des chocs négatifs. A cet effet, le pardon semble garantir une certaine résistance à des événements négatifs potentiels.

D'une certaine façon, la capacité des agents à pardonner peut s'envisager comme une aptitude à résister aux conséquences émotionnelles liées aux événements négatifs pouvant survenir au cours d'une relation (comme les trahisons, les abus, les querelles etc.). La gratitude, en tant que déterminant de la proximité entre les agents, de l'engagement et de la satisfaction au sein d'une relation, est alors d'une importance évidente dans la potentialité de son maintien. Nous soutenons que l'historique des événements positifs (alimenté par la gratitude), entre deux partenaires, améliore la capacité de résistance émotionnelle des individus, face à des événements négatifs potentiels à venir, ce qui contribue ainsi à protéger leur relation. Ici, le rôle de la gratitude, dans la protection des relations, est encore activé de manière préventive, bien que l'action de guérison, réalisée par le pardon, traite, *a posteriori*, l'apparition effective d'événements interpersonnels négatifs et leurs conséquences.

2. La Gratitude et le Côté Radieux des Emotions Morales

De plus, selon Young (2006), la gratitude peut jouer un rôle de soutien de la relation d'une autre manière, à savoir, grâce à son impact sur le maintien de la confiance entre les partenaires, ce qui est cruciallement nécessaire pour permettre un nouveau départ de la relation. Ainsi, un autre élément du rôle protecteur de la gratitude est sa contribution à la conservation de la confiance entre les partenaires, dans la perspective d'accroître les chances que les relations repartent sur une base solide après un choc négatif.

Pour résumer, nous estimons que la gratitude contribue à garantir une certaine résistance des relations face aux événements négatifs, tant par son influence sur le niveau de patience des agents, qu'en vertu de son impact sur la capacité des individus à pardonner, ou par le maintien de la confiance entre les partenaires.

Plus précisément, la gratitude, comme un moyen de réduire l'impatience économique, inhibe les comportements qui menacent les relations. Au contraire, la gratitude favorise l'adoption de choix stratégiques qui permettent la stabilité et la durée d'une relation, même lorsque ces stratégies sont moins attirantes à court terme.

De plus, nous soutenons que la gratitude, en tant que déterminant de la qualité des relations interpersonnelles, est un facteur décisif pour la guérison de celles qui ont été endommagées. Ce processus se réalise grâce à la capacité de la gratitude à favoriser la proximité, l'engagement et la satisfaction au sein d'une relation, ce qui pousse finalement les partenaires à se pardonner les uns les autres, ainsi que, sans doute, à être indulgent (dans le sens de consentir à s'accomoder avec des comportements abusifs) les uns envers les autres, afin d'éviter le sacrifice potentiel d'un lien social précieux.

En outre, Isen et Daubman (1984) ont lié les émotions positives à la flexibilité et à l'inclusion. Ces résultats semblent indiquer que lorsque les individus ressentent des émotions positives, comme la gratitude, ils démontrent une tendance à accepter les gens qui sont différents. Nous pensons que ce phénomène peut s'étendre aux personnes qui présentent des comportements inappropriés (des comportements qui *diffèrent* de la *norme*). De cette façon, il se pourrait que les gens qui sont sous l'influence de la gratitude soient susceptibles d'être plus indulgents et flexibles avec des agents qui adoptent des comportements singuliers, ainsi que d'être prêts à les inclure dans leur groupe malgré tout. En effet, Clark et Waddell (1985) expliquent, par exemple, que le non respect de la norme de réciprocité par un partenaire, dans une relation d'échange, rend cette personne plus abusive et moins attrayante aux yeux de l'autre partenaire. Néanmoins, le même comportement ne provoque pas le même changement de perception lorsque la relation est définie comme étant intime. Nous pensons que ce résultat peut provenir du fait qu'une relation intime est plus susceptible d'être caractérisée par la proximité, l'engagement et la satisfaction, ce qui permet aux partenaires d'être indulgents les uns avec les autres et de rester en bons termes, malgré la violation de la norme.

Par ailleurs, le pardon a également été lié au bien-être mental et physique dans plusieurs études, y compris les travaux de McCullough (2000), qui explique cette association comme découlant du fait que le pardon dans le fait aide les individus à maintenir un ensemble de relations interpersonnelles stables, sur lesquelles ils s'appuient. Par conséquent, la gratitude, en tant que déterminant de la proximité, de l'engagement et de la satisfaction au sein d'une relation, soit en tant que déterminant de second ordre de la capacité des

partenaires à se pardonner, peut être liée au bien-être mental et physique des agents. Les études menées par McCullough et Root (2008) ont tendance à valider ce postulat. En effet, ils ont constaté que l'augmentation du pardon est liée à l'augmentation du bien-être psychologique et, notamment, que le pardon est plus fortement lié au bien-être pour les personnes qui ont déclaré être plus proches de leurs partenaires avant la transgression. Pris dans leur ensemble, ces éléments peuvent expliquer en partie le puissant effet protecteur de la gratitude contre le mal-être (solitude et dépression) suggéré par Chen, Chen et Tsai (2012).

De plus, comme indiqué plus haut, la gratitude est non seulement impliquée dans la protection des relations précieuses (par le biais de la proximité et du pardon), mais elle est encore plus nécessaire pour construire et développer des amitiés durables et des liens sociaux (quelles que soient les menaces auxquelles ils peuvent faire face), qui forment les ressources sociales qui sont essentielles pour la santé, le bien-être et même la longévité (Frederickson, Cohn, Coffey, Pek et Finkel, 2008 ; Algoe, 2012).

Ces derniers éléments nous permettent de faire la transition avec la section suivante dans laquelle nous nous intéressons au lien existant entre la gratitude et le bien-être des individus et nous analysons en quoi ce lien peut expliquer la performance de leur comportement, lorsqu'il s'agit de répondre aux attentes sociales normatives.

2.3. La performance par le bien-être et par la promotion morale affective des comportements coopératifs

2.3.1. Gratitude et bien-être

2.3.1.1. Une corrélation bien identifiée

De nombreuses études empiriques ont mis en évidence le fait que la gratitude est liée à la santé mentale, à la satisfaction dans la vie, au bonheur et au bien-être des individus (par exemple McCullough, Tsang et Emmons, 2004 ; Wood, Maltby, Gillet, Linley et Joseph, 2008 ; Wood, Joseph et Maltby, 2009 ; Chen et Kee, 2008 ; Froh, Kashdan, Ozimkowski et Miller, 2009 ; consulter Wood, Brown et Maltby, 2010, pour une revue sur la gratitude et le bien-être.). De plus, Park, Peterson et Seligman (2004) ont testé l'association entre différentes forces de caractère (de l'anglais « Character strength », notre traduction) et la satisfaction d'une personne à l'égard de sa vie⁵ et ils ont pu montrer, entre autres, que l'espoir, l'enthousiasme, l'amour, la curiosité et la gratitude sont substantiellement liées à la satisfaction dans la vie alors que les autres forces de caractère étudiées ne sont que faiblement associées au bien-être. En conséquence, en plus d'être utile au bon déroulement des interactions sociales, la gratitude se révèle également être l'une des forces de caractère la plus liée à la satisfaction dans la vie.

5. Précisément, ils ont étudié toutes les forces de caractère appartenant à la classification VIA (*i.e.* « Values in actions »). Cette classification répertorie plus de vingt forces de caractère parmi lesquels on retrouve le courage, la citoyenneté, la créativité, la curiosité, l'espoir, l'humour, l'intégrité, l'amour, la modestie ou la gratitude.

2. La Gratitude et le Côté Radieux des Emotions Morales

Plusieurs raisons ont été avancées pour expliquer l'effet bénéfique de la gratitude sur le bien-être. Des recherches en psychologie ont rapporté que les personnes reconnaissantes ont tendance à développer une vision positive du monde, ce qui les empêche de considérer les bienfaits de la vie comme acquis, ainsi que de trop s'habituer à des environnements positifs, ce qui leur permet d'être conscients et de mesurer l'ampleur des bonnes choses qui leur arrivent (McCullough *et al.*, 2001 ; McCullough, Emmons & Tsang, 2002 ; Chen, Wu et Chen, 2015). En outre, la gratitude aide les individus à faire face aux événements négatifs qui surviennent dans leur vie, à travers une reformulation et une redéfinition de ces événements, de sorte qu'ils sont en mesure de trouver un sens positif à la souffrance qu'ils éprouvent (McCullough *et al.*, 2001 ; Lambert, Graham, Fincham & Stillman, 2009 ; Lambert, Fincham & Stillman, 2012 ; Chen *et al.*, 2015). Enfin, quand un individu est le bénéficiaire d'une action prosociale, il peut se sentir choyé, socialement valorisé, estimé, aimé et soutenu, ce qui nourrit aussi son bien-être (Wood *et al.*, 2008 ; Chen *et al.*, 2012 ; Chen *et al.*, 2015).

Cependant, il semble que la pleine réalisation de l'action de la gratitude sur le bien-être des individus soit conditionnée par la possibilité de pouvoir l'exprimer (Watkins, 2004 ; Froh, Sefick et Emmons, 2008). En effet, l'expression de la gratitude est, semble-t-il, capable d'intensifier un état émotionnel positif déjà présent (Froh, Sefick et Emmons, 2008) et Watkins (2004) va même jusqu'à dire que notre satisfaction semble rester incomplète tant que des louanges ou de la gratitude ne sont pas exprimées à la source de notre plaisir. Par ailleurs, d'autres recherches ont mis en évidence la nature désagréable des expressions d'ingratitude (Stein, 1989 ; Suls, Witenberg et Gutkin, 1981), ce qui tend à indiquer que l'absence d'expression du sentiment de gratitude pourrait également être préjudiciable au bien-être des individus qui agissent de manière prosociale.

2.3.1.2. L'expression nécessaire de la gratitude

Ainsi, le manque d'expressivité des agents semble représenter une limite potentielle, bornant la corrélation entre la gratitude et son impact sur la satisfaction dans la vie. Dans la littérature, on retrouve des travaux qui abondent dans ce sens comme, par exemple, ceux de Chen *et al.* (2012) et Chen *et al.* (2015) traitent du rôle modérateur de l'ambivalence dans l'expression émotionnelle (de l'anglais « ambivalence over emotional expression », notre traduction, que l'on note par la suite AEE). L'AEE se caractérise par un déficit interactionnel dans les relations interpersonnelles, qui conduit les individus ambivalents à être soit inexpressifs, car ils inhibent leur désir de s'exprimer, soit à être expressifs, mais à regretter par la suite leur expressivité (Heisel et Montgrain, 2004 ; King, 1998).

Chen *et al.* (2012) explique que l'effet modérateur de l'AEE peut être compris de deux manières.

Premièrement, un déficit dans l'expression de la gratitude peut être due à une mauvaise interprétation de l'interaction sociale comme, par exemple, dans le cas où le destinataire d'un geste bienveillant est incapable de reconnaître pleinement ce que le bienfaiteur a fait pour lui et omet finalement de le remercier. En retour, ce comportement ingrat peut entraîner une probabilité plus faible pour le destinataire d'être à nouveau aidé par le bienfaiteur, ce qui contribue à limiter ses ressources sociales et, en fin de compte, agit négativement sur son bien-être. Sans aller jusque là, le simple fait de ne pas reconnaître ce que les autres font pour soi-même peut inhiber l'émergence de gratitude et de ses bien-

faits en ce qui concerne la santé et le bien-être.

Nous pensons que des cas similaires de mauvaise interprétation au cours d'une interaction sociale peuvent également se produire avec des personnes qui ne souffrent pas d'AEE. Par exemple, deux personnes peuvent avoir des divergences d'opinions sur la façon dont on doit se comporter dans une situation particulière. En effet, les normes sociales et morales sont présentes dans la plupart des interactions sociales et les individus peuvent avoir des différences plus ou moins prononcées dans leur appréciation de ces normes. L'argument que nous souhaitons faire valoir ici est qu'il peut y avoir des désaccords dans la dimension prosociale accordée à un comportement particulier. Ainsi, une personne peut très bien être l'objet d'un acte prosocial de la part d'un autre agent, tout en considérant la réalisation de cet acte comme étant la moindre des choses à faire dans cette situation particulière. Si tel est le cas, bénéficiaire d'un comportement (relativement) bienveillant pourrait ne pas être suffisant pour déclencher l'émergence de la gratitude chez le destinataire de ce geste, ni pour permettre à son influence positive de faire son œuvre. En ce sens, Wood *et al.* (2008) suggèrent que les individus qui ont une forte disposition à ressentir la gratitude ont une vision particulière de la vie qui les poussent à voir l'aide que l'on leur apporte comme plus coûteuse, plus précieuse et plus altruiste. À l'inverse, ils expliquent également que les individus pouvant être qualifiés d'ingrats ont probablement une vision diamétralement opposée.

Néanmoins, on peut penser que l'ingratitude irradiant d'un individu qui peine à reconnaître l'effort entrepris par un bienfaiteur, conduirait ce dernier à ressentir de la (juste) colère, puis éventuellement à se plaindre de ce mauvais comportement. En retour, cela pourrait déclencher des sentiments de culpabilité chez l'individu ingrat, qui pourrait, en conséquence, être désireux de réévaluer la situation en question, ainsi que son opinion sur les attentes qu'il doit avoir vis-à-vis du comportement des autres dans une telle situation. Finalement, après cette mise à jour, l'ancien bénéficiaire ingrat peut devenir plus enclin à ressentir et à exprimer la gratitude, à l'avenir, suite à un même comportement prosocial. Cet exemple hypothétique nous permet d'illustrer comment les émotions morales peuvent agir de concert pour faciliter les relations sociales et leur évolution dynamique. De toute évidence, il existe d'autres cas possibles où la colère du bienfaiteur initial se traduirait par des actions vengeresses pouvant conduire à une escalade de la violence. Ainsi, nous soutenons que la résolution de ces conflits interactionnels pourrait dépendre de la capacité des individus à pouvoir se remettre en question, ainsi qu'à pouvoir remettre en cause leur système de valeurs. Dans cette perspective, les émotions des êtres humains (en particulier les émotions morales) semblent aider les personnes dans la réalisation de compromis nécessaires pour assurer le bon fonctionnement des interactions sociales.

Deuxièmement, Chen *et al.* (2012) souligne que le rôle modérateur de l'AEE peut se matérialiser à travers l'adoption de comportements verbaux et non verbaux inappropriés, en réponse à des actes prosociaux. En particulier, ils expliquent que « l'on peut imaginer que des individus reconnaissants avec une forte ambivalence dans l'expression émotionnelle peuvent être moins susceptibles d'exhiber un comportement approprié en réponse à un bienfaiteur, tel que le fait de sourire ou de dire merci lorsqu'ils reçoivent de l'aide » (Chen *et al.*, 2012, p. 384, notre traduction).

2. La Gratitude et le Côté Radieux des Emotions Morales

Ici encore, nous pensons que des comportements incohérents peuvent être l'œuvre d'individus standards, car des moyens appropriés permettant de communiquer l'expression de la gratitude peuvent ne pas être toujours, ou au moins pas complètement, disponibles dans certains cas. Par exemple, lorsque l'on reçoit un cadeau commun, pour lequel on ne connaît pas la contribution de chacun des bienfaiteurs, il peut s'avérer difficile d'exprimer sincèrement sa gratitude envers chacun d'eux, ce qui pourrait agacer ceux qui ont participé le plus.

Être conscient de telles limites au fonctionnement de la gratitude est vraiment important afin d'embrasser pleinement son influence dans des situations économiques telles que les dilemmes sociaux.

Quoi qu'il en soit, lorsqu'elle n'est pas entravée par les limites mentionnées ci-dessus, l'association substantielle observée entre la gratitude et le bien-être, associée à l'aptitude de la gratitude à promouvoir la coopération, a tendance à rendre attrayante l'idée de provoquer cette émotion spécifique à des fins d'incitation. En comparaison, les émotions morales négatives comme la juste colère, la culpabilité, la honte ou l'embarras, sont désagréables et ont tendance à être corrélées négativement au bien-être des individus. Par exemple, la honte a été clairement identifiée comme étant négativement corrélée au bien-être des agents (Tangney *et al.*, 2007). De toute évidence, ces émotions négatives sont également nécessaires pour soutenir la coopération, grâce à leur rôle dans la mise en place de sanctions morales. Cependant, l'émergence et l'expression de ces émotions négatives englobent généralement un risque d'escalade de la violence (Hopfensitz et Reuben, 2009).

À cet égard, Ryan et Deci (2000) soulignent l'importance de déterminer les conditions qui favorisent, plutôt que celles qui affaiblissent, le potentiel humain, puisque cela peut contribuer, non seulement à la connaissance des déterminants du comportement humain, mais aussi à la conception d'environnements sociaux qui optimisent le développement des individus, leur performance, ainsi que leur bien-être.

En ce sens, dans le contexte des dilemmes sociaux, il semble que favoriser l'induction et l'expression de la gratitude, comme un levier de la promotion de la coopération dans les groupes, soit une stratégie plus attractive, en termes de bien-être, que l'induction d'émotions morales négatives. En d'autres termes, lorsqu'elles sont confrontées à la gestion de situations où la coopération est nécessaire, les institutions en charge (comme les entreprises au moment de décider de leurs politiques de gestion ou les équipes managériales) sont plus à même de favoriser le bien-être social en mettant en œuvre des mécanismes incitatifs qui favorisent l'induction et l'expression d'émotions positives comme la gratitude (comme ceux qui englobent des possibilités de récompense), plutôt que par le choix de mécanismes qui entraînent l'émergence d'émotions morales négatives (comme ceux qui, par exemple, offrent des possibilités de punition).

2.3.2. Le bien-être est-il décisif dans la performance ?

Néanmoins, dans de nombreux cas, on pourrait penser que la mise en œuvre effective de la coopération est plus décisive que le moyen choisi pour l'atteindre. Ainsi, on pourrait juger que c'est une préoccupation secondaire que de s'assurer de promouvoir le bien-être des personnes lorsque l'on cherche à encourager la coopération. Par exemple, on peut se focaliser sur le fait qu'un bien public soit entièrement financé par des contributions, tout en n'accordant que peu d'attention au fait que cela puisse se faire au prix de querelles, plus ou moins douloureuses, entre les partenaires impliqués. Dans le même ordre d'idée, il peut sembler naturel que l'on attende en premier lieu d'une campagne de bienfaisance qu'elle recueille efficacement des fonds, quels que soient les moyens utilisés par les collecteurs qu'elle emploie (comme, par exemple, en jouant sur les sentiments de culpabilité des individus pour provoquer les dons). De même, au niveau organisationnel, certains pourraient faire valoir que le bonheur des employés est moins primordial que leur efficacité au travail.

Cependant, il a été démontré que les personnes qui sont satisfaites dans leur vie sont efficaces dans la résolution des problèmes, démontrent une meilleure performance au travail, ont tendance à être plus résistants au stress, et font preuve d'un meilleur état de santé physique (Veenhoven, 1989 ; Frisch, 2000). Dans une autre étude, impliquant des étudiants de MBA, Wright et Cropanzano (2004) ont également montré que les participants avec fort niveau de bien-être sont de meilleurs décideurs, démontrent de meilleurs comportements interpersonnels et obtiennent de meilleures évaluations globales de performance, par rapport aux sujets qui ont un faible niveau de bien-être.

De cette façon, opter pour des décisions de gestion qui favorisent les émotions positives, comme la gratitude, peut être payant à deux niveaux dans le but d'obtenir de l'effort de la part des agents. D'une part, c'est efficace en tant que moyen de promouvoir la coopération et la réciprocité positive entre les agents et, d'autre part, c'est efficace en tant que moyen d'augmenter le niveau de performance des agents, en raison d'une plus grande satisfaction dans leur vie.

2.3.3. L'influence de la gratitude sur la motivation, un élément clé de la performance et du bien-être

Par ailleurs, Grant et Wrzesniewski (2010) ont démontré que parmi les employés présentant une auto-évaluation méliorative⁶ *seuls ceux* qui valorisent et se sentent concernés par le bien-être des autres, sont en mesure d'attester d'un niveau tangible d'efficacité au travail. Sur la base de cette constatation, ils suggèrent qu'une auto-évaluation méliorative peut ne pas être suffisante pour assurer la performance au travail, puisque des *motivations prosociales* pourraient être nécessaires pour que les employés démontrent des efforts substantiels au travail. La raison qui est donnée pour expliquer ce résultat est que les employés qui sont orientés vers les autres sont probablement plus enclin à anticiper le fait de se sentir coupable de décevoir les autres, lorsqu'ils échouent, et à anticiper l'émergence de la gratitude chez les autres lorsqu'ils réussissent (Grant et Wrzesniewski, 2010).

6. Le terme « auto-évaluation méliorative » signifie que les sujets ont tendance à avoir une haute estime de soi, à se représenter comme hautement efficaces, à avoir des croyances accrues dans leur capacité de contrôle personnel, et à faire preuve d'une grande stabilité émotionnelle (Juge, Locke, et Durham, 1997)

2. La Gratitude et le Côté Radieux des Emotions Morales

Cette dernière interprétation souligne le rôle de la gratitude comme une émotion agréable et que les agents peuvent anticiper comme telle, les motivant alors à fournir des niveaux substantiels d'efforts dans leur travail. L'anticipation de l'émergence de la gratitude chez les autres les aide à réaliser à quel point ces comportements prosociaux peuvent être socialement valorisés, ce qui leur offre une *motivation émotionnelle extrinsèque* à agir de manière prosociale et coopérative. De plus, quand la gratitude anticipée chez le partenaire est effectivement ressentie, l'agent qui est la cible de l'expression de la gratitude, reçoit le signal que son comportement était à la fois utile et souhaité par celui qui l'a exprimée, ce qui tend à promouvoir son estime de soi et peut conduire à l'émergence de l'émotion positive de fierté.

2.3.3.1. Donner du sens, pour mieux s'engager

Une autre façon intéressante par laquelle la gratitude peut avoir un impact à la fois sur le bien-être et sur la performance des agents vis-à-vis ce qui est attendu d'eux apparaît dans l'étude de Flinchbaugh, Moore, Chang et May (2011). En effet, les auteurs ont constaté, chez des étudiants en classe de premier cycle, que la gratitude est corrélée positivement avec le fait de donner du sens à leur activité en classe, ainsi qu'avec leur niveau de participation.

Donner du sens à son travail peut être défini, dans la littérature en gestion, comme le fait de valoriser ses objectifs de travail, au regard de ses propres idéaux ou standards (Flinchbaugh *et al.*, 2011). En outre, les personnes qui ressentent que les rôles qu'ils occupent, dans le cadre de leur activité, sont sensés, rapportent notamment se sentir utiles (Khan, 1990). De son côté, l'engagement porte sur l'implication physique, cognitive et émotionnelle que les individus démontrent quand ils décident de s'investir dans leur rôle au sein d'une activité (Khan, 1990). Il est également à noter que la recherche d'un sens à son activité professionnelle a été identifiée comme l'un des principaux moteurs de l'engagement au travail (May, Gilson et Harter, 2004).

Il semble donc évident que le sens que l'on est capable de donner à sa vie au sein d'une activité, ainsi que le niveau d'engagement que l'on montre envers son rôle dans cette activité, sont des facteurs très importants de l'efficacité dont on fait preuve dans ce rôle. En reliant le fait de confiner ses sentiments de gratitude dans un journal avec le fait de donner un sens à sa présence et de s'engager à participer en classe, Flinchbaugh *et al.* (2011) ont découvert un autre canal à travers lequel la gratitude peut être connectée à la fois à l'épanouissement personnel et à la performance dans une activité. De plus, il semble que la performance soit ici sous-tendue par des motivations qui peuvent être considérées comme *intrinsèques*. En effet, quand un vrai sens est trouvé à une activité, cela peut être suffisant pour que l'on se livre à des efforts considérables pour devenir efficace dans celle-ci, même lorsqu'il n'y a aucune perspective de gains supplémentaires, de quelque nature, que ce soit sous la forme de récompenses matérielles ou sous la forme de plaisirs émotionnels égoïstes.

2.3.3.2. La gratitude dissimulée dans les méandres de la motivation

Les derniers points que nous venons d'examiner soulignent un aspect crucial en ce qui concerne le niveau de performance que les gens sont en mesure de démontrer dans les activités qu'ils exercent : leur motivation. Pendant des décennies, les psychologues

Richard M. Ryan et Edward L. Deci ont sans relâche examiné la motivation humaine et ses déterminants, ce qui les a amené à proposer la théorie de l'autodétermination (Ryan et Deci, 2000).

2.3.3.2.1. La théorie de l'autodétermination

Cette théorie conceptualise notamment l'autodétermination comme un continuum de différents types de motivation, rangés en fonction du degré avec lequel la motivation émane de soi. Plus précisément, ce continuum va de l'amotivation, état de manque d'intention d'agir, à la motivation intrinsèque, qui se réfère à l'accomplissement d'une activité pour la satisfaction inhérente à celle-ci. Entre ces deux extrémités, se trouvent plusieurs formes de motivation extrinsèques, que nous jugeons être les types les plus intéressants de l'autodétermination, en ce qui concerne à la fois les activités professionnelles et des situations de dilemme social. En effet, nous pensons qu'il est assez rare de trouver des employés qui soient véritablement passionnés par leur activité professionnelle. D'autant plus que les emplois sont généralement composés de plusieurs activités, ce qui augmente la probabilité que les gens soient tenus de faire des choses qu'ils ne trouvent pas intrinsèquement intéressantes. Par ailleurs, les dilemmes sociaux sont des situations dans lesquelles les individus ont la possibilité d'agir pour le bien de la communauté, ce qui est loin d'être nécessairement la source de motivation intrinsèque pour les agents économiques.

En conséquence, nous nous concentrons sur la motivation extrinsèque, qui est présente sous diverses formes. Par ordre croissant d'autonomie, nous trouvons d'abord la *régulation externe*, qui concerne les comportements qui sont effectués dans le but de satisfaire une exigence externe ou dans la recherche de récompenses contingentes. Deuxièmement, la motivation peut aussi être le résultat de la *régulation assimilée*, auquel cas les comportements sont effectués soit pour éviter la douleur émotionnelle liée à la culpabilité (ou l'anxiété), soit afin de bénéficier d'un renforcement de l'ego, tel que par l'émergence du sentiment de fierté. Ensuite, la *régulation identifiée* se réfère à l'évaluation consciente d'un objectif de comportement, de manière à ce que l'acte en question soit accepté comme personnellement importante. Finalement, la *régulation intégrée* se produit lorsqu'une *régulation identifiée* est entièrement assimilée au sein de soi, dans le sens où elle a été évaluée et mise en harmonie avec ses autres valeurs dans la vie. Selon Ryan et Deci (2000), les actions entrant dans le champ d'application de la motivation intégrée sont proches de celles qui relèvent de la motivation intrinsèque, sauf qu'au lieu d'être réalisées pour leur plaisir inhérent, elles le sont plutôt dans le but d'obtenir d'autres résultats, dissociables de ce plaisir.

La littérature sur le sujet suggère également que plus la motivation des individus à entreprendre une activité émane d'eux-mêmes, plus ils sont efficaces, persistants et créatifs dans l'exécution de cette activité (par exemple, Deci et Ryan, 1991; Sheldon, Ryan, Rawsthorne et Ilardi, 1997). Compte tenu de cela, la recherche de moyens pour faciliter l'intégration de la motivation extrinsèque est d'un intérêt crucial, s'agissant de la gestion des individus, en particulier dans les situations de dilemme social. Ryan et Deci (2000) ont identifié trois principaux facteurs susceptibles de favoriser cette intériorisation.

2.3.3.2.2. Les déterminants de l'intégration des motivations extrinsèques

Les liens affectifs

Premièrement, Ryan et Deci (2000, p. 73, notre traduction) rapportent que « puisque les comportements extrinsèquement motivés ne sont généralement pas intéressants, la raison fondamentale pour laquelle les individus exécutent initialement de telles actions est parce que ces comportements sont promus, estimés ou valorisés par des proches, vis-à-vis desquels ils se sentent (ou veulent se sentir) attachés ou reliés. Ceci suggère que les liens affectifs, le besoin d'avoir un sentiment d'appartenance sociale et de développer une connexion avec les autres, sont d'une importance capitale pour l'intériorisation ».

En effet, lorsque des personnes importantes à nos yeux comptent sur nous pour répondre à certaines exigences, on peut être plus enclin à prendre ces exigences à cœur, dans la mesure où l'on est conscient que ces comportements peuvent affecter les personnes dont on se soucie. Il est intéressant de noter que les trois adjectifs utilisés par les auteurs, qui sont relatifs au fait de se sentir lié à quelqu'un, d'avoir un sentiment d'appartenance envers un groupe et de se sentir connecté aux autres, ne sont pas sans rappeler ce que nous avons désigné jusqu'ici comme la proximité entre les agents.

De ce point de vue, on peut facilement discerner une première façon par laquelle la gratitude est susceptible d'avoir une influence positive sur l'intégration de la motivation extrinsèque. En effet, comme vu précédemment, la gratitude agit en faveur de la promotion de la proximité entre les agents dans les relations dyadiques, ainsi qu'au sein des groupes. Par conséquent, cela fait de la gratitude un précieux déterminant de second ordre de l'intériorisation de la motivation extrinsèque.

Avant de passer au deuxième facteur facilitant l'intégration de la motivation extrinsèque, il convient de noter que l'influence de la gratitude sur le premier facteur fait écho au concept d'identification organisationnelle.

Comme l'ont indiqué Dutton, Dukerich et Harquail (1994), le niveau d'identification organisationnelle décrit le degré avec lequel les individus parviennent à voir le groupe / l'organisation en tant que partie d'eux-mêmes, dans le cadre de leur processus d'engagement. En outre, cette aptitude est susceptible d'être influencée par les émotions positives comme la gratitude, comme cela est notamment suggéré par Vacharkulksemsuk *et al.* (2011), qui indiquent que les émotions positives contribuent à l'expansion du concept de soi, et ainsi que les expériences associées à la gratitude, ou à d'autres émotions positives, sont censées accroître l'identification organisationnelle.

Vacharkulksemsuk *et al.*, 2011 indiquent également qu'un climat émotionnel positif, au sein des organisations, encourage une augmentation de la force relationnelle des organisations, qui se réfère aux liens et connexions créés et solidifiés, entre les individus, dans les organisations. Enfin, ils affirment que l'identification organisationnelle et la force relationnelle des organisations, qui trouvent toutes deux leurs racines dans la mise en place d'un climat émotionnel positif, contribuent à la croissance et la performance organisationnelle et, par conséquent, à la croissance et au développement de la société⁷.

7. En outre, des travaux ont également montré que les organisations dans lesquelles les employés affichent des niveaux plus élevés d'engagement et expérimentent plus fréquemment des émotions positives, obtiennent une meilleure fidélisation de la clientèle, un renouvellement moins fréquent du personnel, des

En conséquence, il semble que la gratitude, ainsi que d'autres émotions positives, peut aider à améliorer la performance des individus dans leur activité au sein d'un groupe ou d'une organisation, grâce à une motivation plus intégrée pour contribuer à l'effort commun, en partie en raison de la capacité d'un climat émotionnel positif à assurer que les individus se soucient de leur groupe / organisation (dans son ensemble) et de chacun de ses membres. Ceci n'est pas sans rappeler certaines pratiques de gestion pouvant avoir lieu dans les entreprises. Par exemple, les gestionnaires encouragent parfois la mise en œuvre de coutumes comme, par exemple, des cérémonies de cadeaux de Noël entre employés. Dans ce genre de cérémonie, les employés sont associés au hasard, par paire, et chaque individu doit offrir un cadeau (d'une valeur prédéterminée) au partenaire qui lui a été désigné. Ce type de pratique est bon un exemple de canal par lequel on peut activer un rapprochement entre les agents à travers l'émergence de sentiments de gratitude. Ces émotions se traduisent ensuite, par un plus grand sentiment d'appartenance au groupe, ainsi que par une plus grande intégration de la motivation, ce qui favorise de bonnes performances au travail. De toute évidence, il ne serait pas surprenant que ce processus soit encore plus efficace lorsqu'il est renouvelé d'année en année, de sorte que les liens entre les individus puissent être créés et renforcés dynamiquement, grâce à l'influence de la gratitude, dans l'ensemble de l'équipe de travail.

La compétence perçue

Deuxièmement, Ryan et Deci (2000) soutiennent que l'intériorisation de la motivation extrinsèque est aussi fonction de la compétence perçue, puisque les individus sont plus susceptibles de se livrer à une activité quand ils sentent qu'ils sont capables de fournir efficacement leur prestation. Ainsi, ce postulat implique que se sentir compétent pour une tâche spécifique est bénéfique pour encourager une motivation plus intégrée dans sa réalisation, ce qui conduit ensuite à une exécution plus efficace de cette tâche. Ici, on peut être tenté de relier ce phénomène à des questions de confiance et d'estime de soi. En particulier, nous pensons que la plupart des gens ont un besoin constant d'être rassurés sur leurs capacités à entreprendre des actions, notamment les plus difficiles, surtout lorsque leurs comportements sont accomplis au cours d'interactions sociales, qui impliquent la présence des autres et de leurs potentiels jugements critiques.

D'où la nécessité de retours d'information positifs concernant la façon dont les individus sont capables d'accomplir les activités et obligations qui parsèment leur vie quotidienne, peut-être même pour ce qui est des plus ordinaires. En ce sens, nous soutenons que la gratitude, à travers son expression, est impliquée dans la fourniture de ces retours. La raison en est que, comme nous l'avons vu plus haut, l'expression de gratitude communique, à celui qui en est la cible, que son comportement est valorisé et apprécié par le destinataire. Le bienfaiteur, une fois remercié, en déduit naturellement que sa performance a été à la hauteur des attentes de l'autre. L'émotion de gratitude, à travers son aspect communicatif, joue très certainement un rôle clé vis-à-vis du niveau de confiance en soi des agents, ainsi que dans la perception qu'ils ont de leurs compétences, s'agissant de la réalisation quotidienne de leurs activités, en particulier au travail. En somme, la gratitude semble agir à nouveau au niveau du bien-être des individus, en ce qu'elle permet un renforcement de leur confiance en soi, à travers son expression. De plus, nous sommes également amenés à convenir d'une nouvelle voie par laquelle la gratitude peut avoir une action positive sur

ventes nettes plus élevées, et dans une certaine mesure, sont plus productives et rentables (Fleming 2000a, 2000b ; Harter, 2000 ; Harter, Schmidt et Hayes, 2002).

la performance des agents, puisque la perception des compétences prédit des motivations à agir plus intégrées et donc une plus grande efficacité.

L'autonomie

Dernier élément mais non des moindres, Ryan et Deci (2000) indiquent que, plus encore que le fait de se sentir lié aux autres et compétent dans son activité, le fait de se sentir autonome est profondément nécessaire pour que les individus internalisent une motivation extrinsèque. Plus précisément, ils expliquent que « pour intégrer une norme, les individus doivent en saisir le sens et doivent ensuite faire la synthèse de ce sens au regard de leurs autres objectifs et valeurs » (Ryan et Deci, 2000, p. 74, notre traduction). Ils défendent notamment cette hypothèse en se référant au travail expérimental de Deci, Eghrari, Patrick et Leone (1994), qui a mis en évidence que le fait de prodiguer une justification sensée à l'accomplissement d'un comportement sans intérêt apparent, combiné avec des procédures organisationnelles permettant aux agents de se sentir autonomes et liés les uns aux autres, peut promouvoir l'internalisation d'un tel comportement.

Ici encore, l'influence de la gratitude s'avère être d'une valeur considérable. Avant tout, nous savons, d'après les éléments probants examinés plus haut, que la gratitude aide les individus à trouver un sens à ce qu'ils font, ce qui peut, en retour, les conduire à envisager leurs activités comme résultant de leurs propres choix autonomes.

De plus, pour que l'on puisse se sentir autonome, on a besoin de sentir que nos partenaires nous font confiance. Ceux-ci doivent être prêts à se mettre dans une position de vulnérabilité vis-à-vis de nous, pour nous permettre d'avoir le contrôle au moment de faire nos choix. Comme mentionné précédemment, de nombreux résultats empiriques montrent que la gratitude améliore le niveau de confiance entre les agents au sein des groupes. Cette émotion paraît donc participer à la construction d'un environnement propice au fait que chaque agent puisse se sentir suffisamment en confiance pour laisser aux autres le contrôle de leurs propres choix. De ce fait, ceci tend à rendre plus intégrée la motivation conduisant à agir conformément aux attentes des autres et, finalement, augmente le désir des agents de se comporter de manière prosociale.

En somme, ces différents résultats suggèrent que la gratitude joue un rôle positif relativement à chacun des trois facteurs identifiés comme favorisant l'intégration des motivations extrinsèques, qui est déterminant de l'efficacité des individus dans leurs activités. Ceci continue de mettre en lumière la puissance de la gratitude à l'égard de la promotion des actes prosociaux, qui sont principalement motivés de manière extrinsèque, du fait même de leur nature.

2.3. La performance par le bien-être et par la promotion morale affective des comportements coopératifs

Pour conclure, tout semble indiquer que les émotions positives, le bien-être et la performance sont profondément liés. En effet, tous les résultats empiriques que nous avons présentés dans cette section mettent en évidence l'efficacité de l'établissement d'un climat émotionnel positif entre les partenaires, notamment sous l'influence de l'émotion de gratitude, pour ce qui est d'aider à assurer le succès d'un projet collaboratif. Par ailleurs, le renforcement de la coopération et de l'efficacité, en partie grâce à des motivations plus intégrées, ne se produit pas au détriment du bien-être des partenaires, bien au contraire. Dans les faits, cultiver des sentiments positifs tend à favoriser le fait que les agents soient satisfaits dans leur vie, tout en leur permettant probablement d'éviter de se sentir mal à l'aise, voire déprimés, à l'idée d'exécuter les tâches qui sont socialement attendues d'eux.

Nous avons délibérément choisi d'illustrer nos commentaires en mettant l'accent sur la performance au travail, afin d'ancrer le débat dans un cadre relativement concret et pour mieux exposer le fait que la gratitude a définitivement un intérêt pratique dans la promotion de la coopération. De plus, les dilemmes sociaux sont au cœur de l'analyse du présent document et partagent des caractéristiques communes avec de nombreuses situations du monde professionnel (comme le travail d'équipe).

Au vu du rôle essentiel de la gratitude dans les interactions sociales, notamment à travers son implication dans la création et le renforcement des relations interpersonnelles (au niveau dyadique et au sein des groupes) ; à travers son influence sur la promotion des actes prosociaux tels que la coopération, et à travers son impact sur le bien-être, l'efficacité ainsi que la motivation des agents ; il est juste de penser que les économistes doivent accorder de la considération à cette émotion et qu'il est nécessaire d'intégrer, avec précision, celle-ci dans la représentation du processus de décision économique des individus, notamment dans le contexte des dilemmes sociaux.

Troisième partie .

La Modélisation des Emotions
Morales dans le Processus de
Décision Economique

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

3.1. Introduction

Traditionnellement, les économistes orthodoxes basent leurs modélisations du comportement des agents économiques sur deux hypothèses : (1) la rationalité et (2) la recherche de l'intérêt individuel (Sen, 2002). Classiquement, l'individu est donc supposé choisir rationnellement ses actions afin d'atteindre son but, celui-ci étant défini comme le fait de promouvoir son propre bien-être individuel (Vanberg V., 2008). De manière plus formelle, on suppose généralement, en théorie des jeux, que les agents économiques cherchent à maximiser une fonction d'utilité qui prend uniquement en compte leur intérêt individuel (égoïste), capté, le plus souvent, sous la forme des gains matériels qu'ils obtiennent individuellement dans un jeu. Néanmoins, l'économie expérimentale a permis de dégager de nombreuses observations venant nuancer cette conception du comportement humain. Ainsi, bien qu'il existe des travaux expérimentaux dont les résultats tendent à valider une telle représentation des agents économiques (comme, par exemple, les données obtenues par Vernon Smith (1962), montrant une convergence des marchés vers l'équilibre compétitif), dans de nombreuses autres situations économiques, comme les dilemmes sociaux, les prédictions obtenues originellement, en théorie des jeux, peinent à trouver une validité sur le plan empirique (voir Fehr et Schmidt, 2006 et Camerer, 2003 pour des revues sur la question).

Face à cette insuffisance, de nouvelles modélisations ont été développées à travers un aménagement des deux hypothèses constituant la base de la conception classique du comportement humain, en économie. C'est l'hypothèse de recherche de l'intérêt individuel qui a été la plus amendée par les économistes, en autorisant à ce que l'utilité des agents ne se limite pas à leurs seuls gains matériels. La démarche est toujours de considérer que l'agent effectue une maximisation de sa fonction d'utilité, toutefois celle-ci peut dorénavant intégrer d'autres facteurs déterminants, identifiés par les économistes, tels que les gains matériels des autres individus avec lesquels l'agent interagit, des normes morales ou sociales qu'il aura intériorisées (comme la réciprocité, l'équité ou la justice), voire certaines émotions congruentes avec ces différentes composantes (comme la honte, la culpabilité, la colère ou la gratitude), sous la forme de coûts ou de gains psychologiques.

Ce chapitre propose une revue, non exhaustive, des travaux de modélisation du processus de décision des individus, qui dépassent la représentation originelle des agents économiques et qui permettent ainsi de mieux rendre compte des comportements humains. Un des enjeux de cette revue est de souligner le lien existant entre ces différentes modélisations et le rôle des émotions morales dans la prise de décision économique. Nous

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

considérons alors, d'une part, les travaux qui offrent une modélisation « implicite » des émotions (morales), dans le sens où leur rôle n'est pas spécifiquement mentionné, ou encore que la référence qui est faite à ce rôle, dans la formalisation, n'est pas orientée vers une ou plusieurs émotions particulières. Le plus souvent, ceci est permis à travers l'intégration du bien-être d'autrui et de l'influence des normes (morales ou sociales), dans la fonction d'utilité des agents. D'autre part, nous nous intéressons également aux modèles qui proposent une intégration explicite et délibérée de l'influence des émotions morales, en les introduisant expressément au sein de la fonction d'utilité. Ce chapitre met également en lumière comment l'intégration de ces différentes composantes permet une meilleure représentation des comportements humains avec l'émergence d'équilibres plus conformes à la réalité empirique.

De plus, nous avons constaté, lors des chapitres précédents, que les normes sociales ou morales, auxquelles les agents se confrontent, au moment de leur prise de décision, peuvent tenir une place centrale dans la modulation de leurs choix économiques. La question de l'intégration formelle du rapport qu'entretiennent les agents avec les normes existantes dans l'économie revêt, en conséquence, une importance cruciale. De même, une autre question déterminante est celle de connaître les processus qui conduisent à l'émergence de tels comportements normatifs, au sein d'une société.

En plus de définir un cadre conditionnant l'émergence des émotions morales, la présence de normes sociales et morales dans nos sociétés pourrait également, en partie, être le résultat de leur influence. En effet, ces émotions pourraient, d'une part, être à l'origine du développement de valeurs et de jugements moraux participant à l'édification des normes, et d'autre part, avoir un rôle dans leurs subsistances au sein de la société. Ainsi, avant de nous intéresser à la prise en compte formelle du rôle des émotions morales, en économie, nous étudions d'abord certains travaux théoriques examinant la question de la formation et du maintien des croyances et des attentes normatives.

3.1.1. L'émergence et le maintien des normes

3.1.1.1. Différentes motivations pour le respect d'une norme

Le premier modèle auquel nous nous intéressons est celui d'Akerlof (1980), qui s'inscrit dans le cadre d'une norme de salaire équitable (typiquement, le smic dans le cas de la France) et qui permet de rendre compte des aspects affectifs et réputationnels qui sont présents dans le rapport qu'entretiennent les agents avec la présence d'une norme.

Ce modèle a pour but d'expliquer comment une norme, telle que celle du salaire équitable, peut continuer d'exister, alors qu'il est dans l'intérêt immédiat de certaines personnes d'y désobéir. Par exemple, certains chômeurs peuvent préférer travailler pour un salaire inférieur au salaire équitable, plutôt que de rester sans travail. De même, certains employeurs souhaitent embaucher les chômeurs à un salaire inférieur au salaire équitable.

L'hypothèse posée par Akerlof (1980) est que le maintien de la norme dépend du nombre de personnes qui sont prêtes à la respecter. Sa modélisation intègre des coûts psychologiques, liés au fait à la violation de la norme en vigueur, qui dépendent du nombre d'individus dont le comportement y est conforme. De plus, la forme fonctionnelle utilisée introduit également des aspects en lien avec la réputation des agents.

Dans la notation utilisée par Akerlof (1980), x représente la proportion d'individus qui obéissent à la norme et μ la proportion d'agents qui croient à la pertinence de celle-ci. Ici, il est intéressant de noter, que x peut être différent de μ , ce qui correspond au fait que l'on peut respecter la norme sans la juger juste, autant que l'on peut ne pas la respecter alors que l'on croit à sa pertinence. L'auteur définit également la fonction d'utilité des agents comme suit :

$$U = V + cR - d^R d^C \bar{C}$$

Où V représente la source d'utilité classiquement prise en compte par les économistes, soit typiquement, les gains matériels de l'agent. Par ailleurs, c est un paramètre positif et R est une fonction, en lien avec la réputation de l'agent, qui est détaillée ci-après. De plus, d^R est une « dummy » qui est égale à 0 si l'agent respecte la norme en vigueur dans sa communauté, et qui vaut 1 lorsqu'il désobéit ; tandis que d^C est une « dummy » qui vaut 1 si l'agent croit à la pertinence et à la justesse de la norme, et qui est égale à 0 sinon. Enfin, \bar{C} est un paramètre positif qui mesure la force émotionnelle intervenant relativement au non respect de la norme en question et dont l'amplitude peut varier selon que celle-ci est profondément ancrée dans la culture des agents ou qu'au contraire, ils ne la considèrent pas comme étant essentielle.

Le dernier terme de la fonction d'utilité d'Akerlof (1980) est comparable à un coût psychologique lié à la honte ou à la culpabilité de ne pas s'être conformé à la norme. L'agent ne souffre de cette « culpabilité » que dans le cas où il ne respecte pas la norme ($d^R = 1$) et que, dans le même temps, il la considère comme juste et pertinente ($d^C = 1$). Dans les autres cas (s'il respecte la norme, ou s'il désobéit mais qu'il ne croit pas à la pertinence de celle-ci), aucune perte d'utilité n'est engendrée par cette partie de l'expression de la fonction d'utilité de l'agent.

Cependant, le non respect de la norme peut influencer négativement sur l'utilité de l'agent, à travers l'impact qu'elle est susceptible d'avoir sur la réputation de celui-ci. Le terme R représente en effet la perte d'utilité issue de la désapprobation sociale qui est dirigée à l'encontre de l'agent, lorsque celui-ci ne réalise pas le comportement recommandé par la norme. Cette réprobation est envisagée en termes de perte de réputation et traduit la difficulté qu'aura l'agent à développer, à l'avenir, des relations économiques avec les autres membres de sa communauté.

A l'inverse de la partie de la fonction d'utilité qui est relative à la « culpabilité » ressentie par les agents, le terme R dépend uniquement du respect, ou non, de la norme, et aucune considération de jugement des individus, à propos du bien fondé de celle-ci, n'intervient. En ce sens, le rôle des jugements moraux, ainsi que l'influence des émotions morales sur l'établissement de ces jugements, semblent en être absent.

Néanmoins, le caractère réprobateur et punitif qui est sous-jacent à cette partie de l'expression n'est pas sans rappeler les comportements coercitifs que peuvent réaliser les témoins de la violation d'une norme, envers la personne qui en est fautive, sur la base des émotions d'indignation, de dégoût ou de déception¹ qu'ils ressentent.

1. Identifié comme étant à l'origine des comportements d'ostracisme (Haidt, 2003).

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Au total, la formulation d'Akerlof (1980) prend en compte un aspect clairement identifié de l'influence des normes morales ou sociales, qui est que les individus peuvent être animés par deux types de motivations les conduisant à respecter une norme : (1) une motivation *intrinsèque* qui est telle que les agents sont disposés à respecter une norme, même lorsqu'il y a peu de chance que, ce faisant, ils puissent bénéficier de gains instrumentaux, d'actions réciproques futures, ou d'un gain réputationnel (Sripada et Stich, 2006) et (2) une motivation *instrumentale* qui incite les agents à respecter les normes car cela représente un moyen de parvenir à réaliser certains buts (susciter des actes de réciprocité positive, améliorer leur réputation etc.), et non parce que l'adoption d'un comportement socialement adapté représente un but, en soi. Cette dernière motivation se rapporte à la fonction R dont Akerlof (1980) retient, en particulier, la forme spécifique suivante :

$$R(d^R, \mu) = -\mu d^R \bar{R} = \begin{cases} 0 & \text{si l'agent obéit à la norme} \\ -\mu \bar{R} & \text{si l'agent désobéit} \end{cases}$$

Dans cette expression, \bar{R} est un paramètre positif qui mesure la force « réputationnelle » de la norme et dont l'amplitude varie également selon que la norme est considérée comme primordiale, ou au contraire secondaire, au sein de la communauté. Avec cette expression, on voit clairement que le coût réputationnel du non respect de la norme dépend du nombre de personnes μ qui la jugent pertinente.

Pour tout type d'agent, la désapprobation pouvant être attribuée à un mauvais comportement, de la part des autres membres de la société peut être envisagée en termes de punition sociale imposée par les agents respectueux de la norme qui, au même titre qu'une punition matérielle, pèse sur le processus de décision.

L'aspect réputationnel qui est introduit par l'auteur rappelle également la question de l'arbitrage qui existe entre le gain d'utilité pouvant être obtenu, à long terme, à travers le respect de la norme et les conséquences positives que cela engendre, au niveau relationnel, et la perte d'utilité immédiate que l'adoption d'un comportement respectueux de la norme implique, lorsque celui-ci va à l'encontre de l'intérêt individuel de l'agent. A cet égard, ce travail s'inscrit dans le prolongement de celui de Frank (1988), qui considère que les émotions sont présentes pour orienter les agents vers un choix stratégique qui favorise leur bien-être à long terme.

Le modèle fonctionne alors de la manière suivante. Au début de la période, une proportion μ d'agents croit à la pertinence de la norme en vigueur. Chaque individu de la communauté cherche à maximiser son utilité et décide, en conséquence, de respecter ou non, la norme en question, à travers son comportement stratégique. Ainsi, la norme est violée par l'agent si le supplément de gain matériel que cela lui rapporte dépasse le(s) coût(s) lié(s) à son non respect.

Le fait de violer la norme est plus difficile s'agissant des agents qui croient en sa pertinence car, en plus de la perte de réputation engendrée par le non respect de celle-ci, ces agents doivent également se confronter aux effets délétères de la « culpabilité ». Lorsque chaque agent a réalisé son choix, la proportion x d'agents respectant la norme peut être calculée. La dynamique du modèle procède alors comme suit :

$$\dot{\mu} = g(\mu, x) = \beta(x - \mu)$$

Par conséquent, la proportion d'agents croyant à la pertinence de la norme augmente dans le cas où il y a plus d'individus qui la respectent que d'individus l'ayant déjà intériorisée. A l'inverse, cette proportion diminue lorsque $x < \mu$. Le paramètre μ intervenant dans la fonction de coût réputationnel lié au non respect de la norme, on comprend comment cette dynamique régit l'évolution de la norme et peut conduire au renforcement de son influence, à son simple maintien, ou au contraire à sa disparition, au sein de la communauté ou de la société.

En résumé, ce modèle propose une représentation de la manière dont l'investissement affectif, dont chaque agent fait preuve vis-à-vis du respect d'une norme particulière (sous la forme de la présence et de l'ampleur du coût psychologique de « culpabilité » que celui-ci peut subir lors d'une violation), participe, conjointement à l'influence de la pression sociale provenant de l'investissement affectif d'autrui, au maintien, ou au contraire à la disparition de cette norme dans la société. Toutefois, l'auteur ne s'intéresse pas, à proprement parler, à l'émergence des normes particulières (comme par exemple, celle qui est au centre de son modèle). De plus celui-ci n'explique pas comment une action particulière, ou un type d'action particulier, peut être retenu comme étant la référence normative au sein d'une communauté. La norme est ainsi considérée comme un comportement de référence, déjà établi dans la société, mais qui peut disparaître si les agents ne se montrent pas suffisamment vigilants face à son respect.

3.1.1.2. Le rôle de l'approbation sociale dans le rapport à la norme

Les travaux d'Holländer (1990) s'intéressent également à l'approbation qui peut être accordée aux individus pour leur comportement prosocial, en choisissant de s'insérer dans le cadre du jeu du bien public. Un des intérêts de cette modélisation est qu'elle propose une explication, de type « main invisible », à l'émergence d'une norme.

Holländer (1990) considère un groupe de n agents identiques possédant une dotation initiale en bien privé notée π . De plus, on note b le nombre d'unités que chaque agent contribue, à partir de sa dotation, à la production d'un bien public. On suppose enfin que les p unités restantes sont consommées de manière privée. Ainsi, on a :

$$p = \pi - b, \quad 0 \leq b \leq \pi \tag{3.1.1}$$

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Le bien public est produit à partir des contributions agrégées et l'auteur suppose que le coût de production de ce bien est proportionnel au nombre d'agents présents dans le groupe, ainsi la quantité de bien public finalement obtenue, à partir de l'ensemble des contributions est :

$$C = \frac{1}{n} \sum_{i=1}^n b_i$$

Par ailleurs, dans ce modèle, la contribution b implique un montant total d'approbation absolu $s(b)$, qui représente la manière dont les autres membres du groupe valorisent un tel niveau de contribution.

Formellement, l'auteur suppose que le niveau d'approbation absolu $s(b)$ ressenti par les agents, peut s'exprimer de la manière suivante :

$$s(b) = (1 - \beta)wb - \beta w(b - c) = w(b - \beta c) \quad (3.1.2)$$

Où w est un paramètre représentant le taux d'approbation.

Le niveau d'approbation absolu ressenti par l'agent dépend ainsi positivement de son niveau de contribution individuel b et négativement du niveau de contribution moyen c .

Pour justifier une telle expression, l'auteur indique que, dans le cas d'un cadeau par exemple, la *gratitude* ressentie vis-à-vis de l'individu qui offre ne dépend pas seulement de la valeur absolue du cadeau, mais aussi de la façon dont la valeur attribuée à ce cadeau se compare à celles de cadeaux similaires.

Bien que l'inspiration de cet aspect de la formalisation soit avant tout intuitif, l'auteur fait clairement référence à la gratitude et Wood, Maltby et Brown (2011) ont plus tard corroboré expérimentalement l'existence de ce type « d'effet de ranking » (ou effet d'ordre) concernant cette émotion morale.

Ainsi, on peut interpréter $s(b)$ comme mesurant l'intensité des émotions positives ressenties par l'agent, du fait de percevoir que son comportement est valorisé par les autres membres du groupe, suite à l'émergence, chez eux, de l'émotion de gratitude. L'auteur justifie l'introduction d'un tel paramètre par le fait que les êtres humains semblent avoir une préférence pour le fait d'être l'objet de l'évaluation positive des autres, ainsi que pour le fait d'être la source de l'émergence d'émotions positives chez eux.

Toutefois, l'auteur explique que les individus, en plus de vouloir être aimés ou admirés, veulent également l'être (aimés et admirés) plus que les autres. Holländer (1990) suppose donc que les individus s'intéressent également à la grandeur $s(b) - s(c)$, où $s(c)$ représente le niveau d'approbation qui est associé au comportement moyen de contribution. Cette

grandeur mesure alors l'approbation relative que l'agent perçoit chez ses partenaires. Finalement, l'auteur définit la variable de mesure de l'approbation, utilisée par les agents, comme étant une somme pondérée de la mesure absolue et de la mesure relative de l'approbation :

$$a = (1 - \alpha)s(b) - \alpha[s(b) - s(c)] = s(b) - \alpha s(c) \quad (3.1.3)$$

Par ailleurs, la fonction d'utilité de l'agent est définie de la manière suivante :

$$u = u_p(p) + u_c(c) + u_a(a)$$

Par conséquent, on suppose que les agents valorisent à la fois leur consommation en bien privé, leur utilisation du bien public et, à travers le terme $u_a(a)$, le niveau d'approbation qu'ils reçoivent, de la part des autres, pour leur contribution au bien public. Ce dernier terme, même s'il n'est pas explicitement présenté comme cela par l'auteur, peut se rapprocher du sentiment de fierté ressenti par les agents lorsqu'ils réalisent une action prosociale (correspondant ici au fait de contribuer conséquemment au bien public), et plus particulièrement lorsque leur action est meilleure que celle qui est faite, en moyenne, dans le groupe.

D'après les équations (3.1.1), (3.1.2) et (3.1.3), la fonction d'utilité peut être réécrite de la manière suivante :

$$u = u_p(\pi - b) + u_c(c) + u_a[w(b - \sigma c)]$$

Où $\sigma = \alpha + \beta$.

Ainsi, étant donné la spécification des arguments de la fonction u_a , on peut observer, selon la forme prise par cette fonction, un niveau d'approbation ressenti qui soit négatif, dans le cas où la contribution individuelle de l'agent est insuffisamment importante, par rapport à la contribution moyenne. Dans un tel cas, l'agent se voit alors exprimer la désapprobation de ses partenaires, que l'on peut interpréter comme étant issue de leurs sentiments d'indignation ou de dégoût. Cette désapprobation entraîne une baisse d'utilité que l'on peut alors rapprocher d'un coût psychologique associé à l'émergence des émotions de honte ou de culpabilité.

Dans le même temps, plus la contribution d'un agent est grande, plus l'utilité qu'il retire de sa consommation en bien privé diminue. L'agent réalise par conséquent, au moment de son choix, un arbitrage entre son plaisir matériel individuel et le plaisir qu'il dérive du statut social qu'il obtient à travers son comportement contributif.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

La résolution du modèle s'effectue à travers la maximisation de l'utilité de l'agent par rapport à son niveau de contribution b . Holländer (1990) démontre qu'il existe un unique équilibre « d'échange social », caractérisé par des contributions individuelles cohérentes avec la contribution moyenne (dans le sens où on a $b = c$, à l'équilibre) et des taux d'approbation individuels v^2 , accordés par les agents, qui sont cohérents avec le taux d'approbation général reçu par chaque agent pour son comportement individuel w (dans le sens où $\nu = w$).

Selon les valeurs des paramètres, certains équilibres sont tels que les contributions individuelles s'établissent au niveau de la contribution moyenne, $b^* = c^*$, ce qui en fait alors une norme de comportement, émergeant à la suite des interactions sociales ayant lieu entre les agents, notamment en termes d'échange d'approbation. Lorsqu'un tel équilibre est implémenté, tous les individus contribuent le même montant, correspondant au niveau moyen de contribution, et qui permet de fournir la quantité optimale de bien public c^* . L'émergence d'un comportement normatif est donc, dans ce modèle, considéré comme étant le simple fruit de la confrontation d'agents économiques rationnels, pourvus d'une préférence pour le fait d'être évalué positivement par leur prochain. Toutefois, on peut questionner le réalisme d'un tel résultat, au sens où il n'est pas capable d'expliquer comment les comportements effectifs des individus peuvent être différents (ne serait-ce que légèrement), à l'équilibre, tout en étant influencés par une même norme.

De plus, ce résultat est partiellement lié à la forme de la fonction u_a (définissant l'utilité qui est issue du niveau d'approbation que reçoit un agent pour son comportement), qui dépend positivement du niveau de contribution b de l'individu et négativement du comportement de contribution moyen c . Par conséquent, le niveau de contribution b doit être suffisamment important, par rapport au comportement moyen, pour que l'individu puisse bénéficier d'un statut social qui le satisfasse (la partie u_a de l'utilité entraîne donc le niveau de contribution à la hausse).

En outre, à partir du moment où l'individu bénéficie d'un haut statut social, celui-ci est poussé à moins contribuer, afin de bénéficier de plus de bien privé (la partie u_p de l'utilité représente une force à la baisse pour le niveau de contribution).

De cette façon, le résultat d'émergence de la norme de contribution dans le jeu du bien public (consistant à contribuer un montant égal à la moyenne de contribution du groupe) semble directement issu des hypothèses du modèle et, en particulier, de la sensibilité de l'approbation qui est perçue par les individus, à la contribution moyenne du groupe. Finalement, le fait que Holländer (1990) pose cette dernière hypothèse revient presque à ce qu'il suppose que le comportement moyen observé dans le groupe correspond à la norme qui est en vigueur au sein de ce groupe.

2. Etant donné que le paramètre v est le taux d'approbation qui est accordé par chaque agent (i) à la contribution d'un partenaire particulier (j). Ce taux dépend de l'utilité que cette contribution rapporterait à l'agent i , si tous les autres agents, à part lui (i), adoptaient le même comportement de contribution.

3.1.1.3. Une recherche rationnelle de la norme

Plus récemment, Brekke, Kverndock et Nyborg (2003) ont considéré une origine personnelle de la norme, au sens où l'agent fixe lui-même la valeur de son comportement normatif, sur la base de ses propres convictions morales. Les auteurs envisagent un mode particulier de détermination de la norme, qu'ils considèrent commun à tous les agents. Il est ainsi supposé que les membres de la société partagent tous une philosophie utilitariste, dans le sens où ils pensent que l'intérêt de chaque individu compte, de la même façon, dans l'évaluation du bien-être social, laquelle prend la forme suivante :

$$W = u_1 + \dots + u_n$$

Où W désigne le bien-être social, n est le nombre d'agents et u_i représente l'utilité de l'agent $i \in (1, n)$.

Sur la base de ce critère, chaque agent i détermine sa contribution morale idéale e_i^* ³, en répondant à la question suivante : « Quelle action maximiserait le bien-être social, si chacun des autres individus adoptait le même comportement que moi (et si chacun des autres individus avait la même perception, sous entendu, était caractérisé par la même forme fonctionnelle) ? ». Ce qui correspond à réaliser la maximisation de la somme des utilités W par rapport à e_i , en considérant que chacun des autres joueurs entreprend ce même niveau d'effort.

Chaque agent emploie la même méthode pour déterminer son idéal moral, lequel peut cependant varier, selon les individus, dès lors que l'on introduit de l'hétérogénéité au niveau des caractéristiques individuelles des agents, en termes de préférences (forme de u) ou de dotations initiales.

L'originalité de cette conceptualisation vient du fait qu'elle tient compte de la part de subjectivité qui demeure dans la manière dont les agents abordent leur relation à la norme. Paradoxalement, cela représente également un écueil, au sens où cette modélisation néglige le caractère profondément collectif que possède une norme morale ou sociale (si ce n'est à travers le fait que chaque agent met en œuvre la même méthodologie pour déterminer le comportement qu'il considère comme normatif).

A cet égard, Fehr et Gächter (2000) définissent une norme sociale comme étant (1) une régularité de comportement (2) basée sur une croyance *partagée* (par les individus constitutifs du groupe) concernant la manière dont les individus doivent se comporter, (3) dont la mise en application déclenche l'utilisation de sanctions informelles à l'encontre de ceux qui ne la respectent pas. Les normes morales ou sociales ont donc, de ce point de vue, une origine partagée et collective, ou tout du moins une capacité à s'imposer, au fur et à mesure, de manière uniforme, à chacun des membres du groupe.

3. Brekke *et al.* (2003) considèrent un jeu du bien public, où la contribution se fait sous la forme d'un effort, mesuré en unité de temps.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

S'il existe une part de subjectivité dans le rapport à la norme, celle-ci doit plutôt se retrouver dans la perception et l'interprétation qui est faite de la norme. Par exemple, si l'on considère que la norme d'équité s'impose entre les individus, on peut se poser la question de savoir ce qu'est l'équité, en fin de compte. Lorsque je fournis plus d'effort que mon partenaire pour obtenir un surplus que l'on doit se partager, le paiement équitable est-il le paiement égalitaire ou bien un paiement proportionnel au niveau d'effort ?

Un autre aspect subjectif, concernant le rapport des individus avec une norme, réside évidemment dans la prise de décision individuelle et le choix final de consentir, ou non, à la respecter.

La réflexion mise en œuvre par les auteurs est séduisante pour les économistes classiques, en ce sens qu'elle théorise l'émergence des normes comme étant le produit d'un processus de questionnement, réalisé en toute « conscience » par l'individu, et qui conduit à une réponse déterminée de manière rationnelle, sur la base d'un critère utilitariste. Cependant, certains travaux de psychologie, comme notamment ceux réalisés par Haidt (2001), remettent en cause la pertinence d'une approche (seulement) rationnelle des jugements moraux, ce qui, par extension, pose un doute sur le réalisme d'une conception de l'établissement des normes, qui dépendrait uniquement de processus cognitifs, sans laisser de place aux émotions des individus.

3.1.1.4. Dynamiques de sélection, imitation et apprentissage

Une possibilité alternative, pour trouver une explication théorique à l'émergence des normes, est alors de s'intéresser au courant de pensée de la théorie des jeux évolutionnistes (TJE). Cette théorie ne s'intéresse pas, à proprement parler, à la détermination de comportements à qui l'on choisit de prêter une résonance normative, sur la base de quelconques critères. Elle considère plutôt la façon dont certains comportements, parmi une multitude d'autres candidats possibles, arrivent à s'imposer au sein d'une population et à devenir des « patterns » comportementaux, pour enfin émerger en tant que norme sociale. D'après Kolstad (2007), les modèles de la TJE questionnant l'émergence des normes sociales se caractérisent principalement par l'étude de situations dans lesquelles : (1) les agents sont regroupés de manière répétée pour participer à des jeux, et (2) il existe un processus dynamique qui décrit la manière dont les joueurs adaptent leurs comportements au fil du temps. Ce processus dynamique s'apparente généralement à un processus de sélection qui rappelle les travaux existant en biologie (théorie darwinienne de la survie du mieux adapté).

Un exemple de processus dynamique de sélection est celui du réplicateur dynamique. Ce processus implique qu'au sein d'une population, la part x_i des individus jouant la stratégie i , croît à un taux qui est égal à la différence entre le paiement escompté de la stratégie i et la moyenne des paiements escomptés offerts par l'ensemble des autres stratégies possibles pour les agents (Kolstad, 2007), soit, formellement :

$$\frac{\dot{x}_i}{x_i} = \pi_i - \pi_{moyen}$$

Ainsi, il est supposé, à travers ce processus, que les meilleures stratégies, en termes de gains matériels escomptés, perdurent et se développent au sein de la population. Il est également supposé qu'au contraire, les stratégies proposant un « rendement » (en termes de gains escomptés) inférieur au « rendement moyen » sont vouées à voir leur utilisation décliner jusqu'à une éventuelle disparition.

Les dynamiques proposées par la TJE, en ce qui concerne l'émergence des comportements normatifs, sont cependant plus sophistiquées que le processus de sélection biologique que nous venons de voir. En effet, cette conception considère généralement que les agents ont une capacité d'imitation et d'apprentissage vis-à-vis de leurs choix stratégiques (Kolstad, 2007), mais également que l'ensemble des stratégies à disposition des agents est susceptible d'évoluer au gré de mutations résultant, par exemple, de l'entrée de nouveaux individus au sein d'un groupe ou bien encore d'erreurs ou d'expérimentations, réalisées par des agents déjà présents.

En somme, dans le cadre de la TJE, l'obtention d'une norme sociale, à l'équilibre de long terme, peut être considérée comme l'émergence d'un comportement spécifique, sélectionné à travers un processus dynamique, lequel consiste en une comparaison, en termes d'utilité, des conséquences de cette action stratégique, avec celles des différentes autres stratégies possibles, pour une population d'agents. Toutefois, cet équilibre est sensible à l'apparition de nouvelles stratégies (lors de mutations ou lors de l'introduction de nouveaux joueurs), lesquelles peuvent potentiellement être sélectionnées, par la suite, afin d'occuper la fonction de norme.

Les modèles « d'apprentissage » offrent une conception alternative de la dynamique régissant l'établissement et le maintien des normes de comportements. Ainsi, Sen et Airiau (2007) envisagent l'émergence et le maintien des normes comme étant le résultat d'un « apprentissage social » effectué, de manière dynamique, par les agents. Les auteurs considèrent une population d'agents participant ensemble à un jeu (les jeux examinés par les auteurs étant souvent à équilibres multiples), de manière répétée, durant un grand nombre (fini) de périodes. A chaque période, deux individus de la population sont sélectionnés, au hasard, pour participer à une répétition du jeu. Une originalité de ce modèle tient donc au fait qu'il repose sur un processus d'apprentissage tel que chaque agent ignore avec quel autre individu il sera sélectionné pour jouer (à la différence d'une modélisation où chaque répétition du jeu s'effectue avec le même joueur).

En outre, les auteurs envisagent trois types d'apprentissage, sous la forme de trois algorithmes qui modulent (différemment) le choix stratégique des agents et qui permettent finalement d'aboutir à une convergence des comportements sur un équilibre normatif. Ces algorithmes d'apprentissage peuvent, par exemple, prendre la forme suivante : un agent essaie d'abord une action, pour un état particulier du jeu, et évalue ensuite les conséquences, en termes de gains ou de pertes immédiates, qui découlent de son choix. En essayant toutes les actions, pour tous les états possibles, de manière répétée, il détermine (apprend) quelles actions sont en mesure de maximiser son utilité escomptée de long terme (Q-learning, Watkins et Dayan, 1992). Sen et Airiau (2007) montrent alors, à l'aide de simulations, que la vitesse de convergence des comportements individuels vers un équilibre normatif, dans des jeux proches du dilemme du prisonnier, est une fonction négative du nombre d'individus composant la population, ainsi que du nombre d'actions possibles.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Par ailleurs, ils montrent que cette vitesse peut également dépendre de l'hétérogénéité des algorithmes d'apprentissage qu'utilisent les agents, au sein de la population.

En résumé, avec cette conceptualisation, la sélection de l'action stratégique, qui émerge en tant que norme, à l'issue du processus de coordination des joueurs, est faite au travers d'un apprentissage s'effectuant par l'expérience (et non plus à travers un processus d'imitation, comme cela peut être le cas dans certains modèles de TJE). Néanmoins, cette sélection se fait toujours sur la base du calcul que représente la comparaison entre les différents niveaux d'utilité relatifs aux multiples comportements possibles pour chaque agent, ce qui ne laisse aucune place à une quelconque analyse des conduites, en termes d'éthique ou de morale.

Ainsi, Orléan (1997) s'oppose à une telle explication de l'émergence des normes sociales ou morales. D'après lui, la faiblesse de ces modélisations tient au fait qu'elles envisagent toute dimension normative comme étant absente de jugements (moraux) individuels. En effet, Orléan (1997, p. 6) indique : « La seule question pertinente est de savoir quelle stratégie produit les gains les plus élevés sans que jamais n'intervienne une évaluation portant sur la pertinence, la justesse ou la conformité des conduites » .

En particulier, les modèles d'apprentissage et de TJE ne permettent pas de rendre compte de l'influence des émotions dans l'établissement des jugements moraux. Ces modèles ne prennent pas non plus en compte le fait que les émotions (morales) jouent sans doute un rôle crucial, d'abord dans l'édification même des normes et ensuite dans la modulation des comportements individuels, de sorte à favoriser le respect de celles-ci, ainsi qu'à pousser les individus à désapprouver profondément les personnes qui violent les règles sociales, ou au contraire à donner leur approbation à ceux qui adoptent une conduite morale exemplaire.

En ce sens, nous avons constaté, lors de cette section, que l'on peut interpréter certaines des modélisations revues de sorte à laisser une place au rôle des émotions, lorsqu'une référence explicite à celles-ci n'est pas directement faite par les auteurs.

Finalement, il n'y a pas de réel consensus, au sein de la littérature, sur la question de l'émergence des normes et certains des modèles que nous avons revu ont probablement plus de pertinence vis-à-vis de la question du maintien d'une norme déjà en place. D'après les psychologues Sripada et Stich (2006), les normes et les sanctions sont présentes de manière universelle dans toutes les sociétés humaines. De plus, même s'il existe quelques différences culturelles dans la manière d'appliquer les normes morales, celles-ci ont tendance à se concentrer sur certains thèmes généraux comme l'interdiction d'assassiner, de voler, de commettre un acte incestueux, ou bien des thèmes comme le partage, l'équité, la réciprocité, l'entraide ou encore la coopération. Ces normes pourraient alors avoir une origine très ancienne et être transmises, acquises et appliquées à travers des processus éducatifs d'apprentissage et d'intériorisation, lesquels fonctionnent vraisemblablement sur la base de l'émergence des émotions, et notamment des émotions morales.

Par exemple, une des normes que l'on retrouve, dans la plupart des modèles de préférences sociales, consiste à définir le gain équitable comme un point focal (Schelling, 1960),

3.2. L'intégration implicite des émotions morales dans les modèles économiques

au regard duquel les agents sont en mesure de juger, aussi bien leurs propres comportements, que ceux des autres joueurs. La norme d'équité, dont le caractère universel n'est plus à démontrer, pourrait provenir d'un sens inné de justice, ayant évolué, chez les êtres humains (ainsi que chez d'autres animaux étant capable d'interagir régulièrement avec des partenaires identifiés, comme certaines espèces de primates), sur la base des émotions morales qui émergent chez les individus (en fonction de son respect ou de son non respect), de sorte à promouvoir la pérennité des relations interpersonnelles.

Le déficit de réponse précise à la question de l'émergence des différentes normes permet peut être d'expliquer en partie pourquoi les modèles de préférences sociales sont souvent considérés comme *ad hoc*, au sens où l'on considère que les conclusions auxquelles ces modèles arrivent sont directement issues des hypothèses posées dans le modèle. En effet, bon nombre des formalisations que nous avons vu et que nous allons continuer de détailler dans ce chapitre, font référence à la présence d'une norme de comportement, autour de laquelle les équilibres dégagés tendent à se concentrer.

Toutefois, les différentes normes morales qui sont présentes dans nos civilisations méritent d'être identifiées et intégrées (même si cela doit se faire de manière exogène) aux modèles économiques puisque, bien que leurs origines ne soient pas clairement expliquées, elles ont indubitablement une influence sur le comportement des agents.

Ainsi, le fait de prendre en compte l'importance des normes morales dans la prise de décision des agents, permet aux modèles économiques de gagner en pertinence et en qualité de prédiction. Ce faisant, ces modèles saisissent des mécanismes psychologiques qui peuvent s'interpréter comme étant le résultat de l'influence des émotions, quand cette influence n'est pas volontairement prise en compte par les auteurs.

Dans la suite de ce chapitre, nous examinons ainsi, dans un premier temps, certains travaux marquants de l'économie comportementale présentant des formalisations intégrant implicitement (ou sans référence spécifique) le rôle des émotions (morales) dans la modulation du choix des agents. Avant, dans un deuxième temps, de nous intéresser aux modèles qui font le choix de prendre en compte, dans leur représentation du processus de décision des agents, l'influence d'une ou plusieurs émotions morales spécifiquement identifiées.

3.2. L'intégration implicite des émotions morales dans les modèles économiques

Comme notre revue tend à le montrer, il semble que le lien existant entre normes morales et émotions morales représente un aspect clé dans la modélisation de la prise de décision économique, pour expliquer le comportement des agents, dans les situations de dilemme social. Pour autant, ce lien n'est pas toujours représenté de manière explicite. En effet, de nombreux modèles ont vocation à intégrer le rôle des normes et de certains mécanismes psychologiques qui sont relatifs aux affects, mais sans les formaliser explicitement, ou tout du moins sans spécifier l'impact des différentes émotions morales sur le comportement des agents. Par conséquent, dans cette section, nous offrons une réinterprétation de certains travaux fondateurs, en économie comportementale, de sorte à découvrir comment ces

modèles sont en mesure d'intégrer le fonctionnement de certaines émotions morales, dans la modulation du choix des agents.

3.2.1. La prise en compte de la norme d'équité

Fehr et Schmidt (1999) présentent un modèle d'aversion à l'iniquité qui entend fournir une justification théorique à la manière dont les individus se comportent dans certains des jeux économiques auxquels des sujets participent en laboratoire. Dans leur modèle, la fonction d'utilité de l'agent dépend à la fois de ses propres gains monétaires dans le jeu, mais également des gains des autres joueurs. Formellement, si on considère n joueurs indexés par $i \in \{1, \dots, n\}$ et que l'on note $x = x_1, \dots, x_n$ le vecteur des gains monétaires des n joueurs, la fonction d'utilité du joueur i est donnée par :

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\} \quad (3.2.1)$$

Pour simplifier, on peut également considérer le cas où $n = 2$:

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}; i \neq j \quad (3.2.2)$$

Le premier terme représente le gain monétaire du joueur i dans le jeu. Les deux termes suivants sont relatifs à l'aversion de l'agent i à l'iniquité potentiellement résultante des décisions stratégiques prises par l'ensemble des joueurs. Si $\alpha_i = \beta_i = 0$ l'utilité du joueur i ne dépend plus que de ses propres gains monétaires, et on retrouve alors la représentation traditionnelle des agents économiques, en théorie des jeux. Ainsi, on voit clairement ici que ce modèle, comme la plupart des modèles de préférences sociales, se construit en prolongement et dans la continuité du modèle économique standard.

A l'inverse, si $\alpha_i \neq 0$ et $\beta_i \neq 0$, alors l'utilité du joueur i s'ouvre aux gains monétaires de l'autre joueur. Il convient de remarquer ici que le fait que les deux paramètres α et β soient indicés, rend compte de la différence de personnalité qui peut exister entre les agents.

Le deuxième terme de la fonction mesure la perte d'utilité provenant d'une inégalité désavantageuse pour l'agent i , tandis que le dernier terme mesure la perte d'utilité issue d'une inégalité à l'avantage de l'agent i . On remarque ici que, pour un niveau de gains matériels x_i fixé, U_i est maximale lorsque les deux derniers termes sont nuls, c'est à dire lorsque $x_i = x_j$, ce qui correspond à une situation où les agents respectent la norme d'équité (en considérant que l'équité est obtenue lorsque chaque joueur obtient un gain monétaire égal, cette règle n'est pour autant pas la seule possible).

Fehr et Schmidt (1999) introduisent donc dans leur fonction d'utilité, une norme de référence qui influe sur le comportement des agents l'ayant intériorisée (ceux pour qui

3.2. L'intégration implicite des émotions morales dans les modèles économiques

$\alpha_i \neq 0$ et $\beta_i \neq 0$). En effet, le non respect de cette norme implique des conséquences négatives, en termes d'utilité, pour l'agent i . L'interprétation des auteurs est la suivante :

lorsque $x_i < x_j$, le résultat du jeu est défavorable, en termes monétaires, au joueur i , et celui-ci, du fait de son aversion à l'iniquité, subit le coût psychologique $\alpha_i(x_j - x_i)$.

Symétriquement, lorsque $x_i > x_j$, le résultat du jeu est favorable au joueur i et cela provoque l'apparition d'un coût psychologique $\beta_i(x_i - x_j)$.

A travers une telle formalisation, les auteurs supposent donc une préférence pour la norme d'équité. Néanmoins, ils n'expliquent pas comment cette préférence survient chez les agents. Les auteurs ne donnent pas non plus d'interprétations explicites, des mécanismes à l'œuvre lorsque cette préférence intervient. A la lumière des recherches réalisées sur les émotions morales, il est possible de trouver certaines similarités entre la modélisation proposée par Fehr et Schmidt (1999) et les faits stylisés dégagés vis-à-vis du rôle de ces émotions dans la prise de décision des agents économiques.

Ainsi, dans ce modèle, lorsque $x_i < x_j$, la norme d'équité n'est pas respectée et le joueur i se sent lésé par rapport à l'autre joueur, ce qui se traduit par l'apparition d'un coût psychologique $\alpha_i(x_j - x_i)$ chez lui.

Dans la littérature, le coefficient α_i a souvent été associé à l'émotion d'envie, qui s'exprime dans les situations où l'on est envieux d'autrui. Il est vrai que la présence et l'influence de cette émotion est pertinente dans cette situation et il est tout à fait possible que seule l'émotion d'envie ait été envisagée par les auteurs au moment de la genèse de leur modèle. Cependant, une interprétation de ce coût psychologique en termes d'émotions morales de condamnation d'autrui a également du sens. Ces émotions, comme l'indignation (colère), le dégoût ou le mépris, émergent dans les situations caractérisées par la violation d'une norme morale ou sociale commise par un autre agent et sont à valence négative, et donc désagréables à ressentir, ce qui peut se traduire par une diminution de l'utilité.

De surcroît, en considérant $\alpha_i > 0$, les auteurs montrent que, dans une telle situation, le joueur i est prêt à sacrifier une unité monétaire si cela réduit les gains de son adversaire de $(1 + \alpha_i)/\alpha_i$. Par conséquent, l'agent i peut être prêt, dans le but de maximiser son utilité, à punir son partenaire, même lorsque la punition est coûteuse. Or, la tendance à l'action des émotions de condamnation d'autrui est justement caractérisée par ce type de comportements punitifs permettant un certain soulagement de l'état affectif (la colère, par exemple, étant une émotion ressentie de manière relativement intense).

Par ailleurs, lorsque $x_i > x_j$, la norme d'équité n'est pas respectée et, cette fois, c'est le joueur j qui est lésé. Le coût psychologique subi par le joueur i peut alors s'interpréter en termes d'émotions morales auto-conscientes à valence négative. Ces émotions, comme la culpabilité, la honte ou l'embarras, interviennent lorsque l'on se rend soi-même responsable de la violation d'une norme morale ou sociale. Elles sont également désagréables à ressentir et leur émergence peut aussi être interprétée comme une baisse de l'utilité, comparable à celle que subit l'agent dans le modèle.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Avec ce premier modèle, on voit comment certaines spécificités des émotions morales peuvent être captées formellement, même lorsqu'elles ne sont pas explicitement identifiées par les auteurs. On retrouve ce phénomène dans de nombreuses autres modélisations, comme celle de Bolton et Ockenfels (2000).

Le travail de Bolton et Ockenfels (2000) est relativement proche de celui de Fehr et Schmidt (1999), mais présente quelques particularités intéressantes, notamment du point de vue de l'information que possèdent les joueurs. Dans ce modèle, comme dans le précédent, l'utilité des agents dépend à la fois de leurs gains dans le jeu, et de la part relative de leurs gains, vis-à-vis du reste des joueurs. Cependant, là où le précédent modèle suppose une comparaison individuelle des gains, joueur par joueur (voir (3.2.1)), Bolton et Ockenfels (2000) supposent une comparaison des paiements du joueur avec le total des gains obtenu par l'ensemble des joueurs.

Ainsi, les joueurs sont ici « sensibles aux parts relatives » dans le gain total, tandis que le modèle de Fehr et Schmidt (1999) suppose que les joueurs sont « sensibles aux écarts absolus de gains » (Garapin et Ruffieux, 2009).

Avec une telle spécification, la dimension collective de la norme est peut-être encore plus affirmée, dans la mesure où le groupe (au sens de l'ensemble des joueurs) est placé au cœur de la formalisation du rapport que les individus entretiennent avec la norme. Cet aspect est intéressant puisque, du point de vue de l'émergence des émotions morales, le niveau individuel (je ressens de la colère à l'encontre de celui qui a réalisé une violation) est au moins aussi pertinent que le niveau collectif (je ressens de la culpabilité vis-à-vis d'un groupe de personnes pour m'être attribué une part plus importante que le gain moyen du groupe).⁴

Les auteurs ne se limitent pas à une forme spécifique de la fonction d'utilité des agents, mais proposent tout de même, dans le cas où le nombre de joueurs n est égal à deux, un exemple de fonction repris ici par souci de clarté :

$$U_i(x_i, \sigma_i) = a_i x_i - \frac{b_i}{2} \left(\sigma_i - \frac{1}{2} \right)^2 \quad (3.2.3)$$

Avec $a_i \geq 0$ et $b_i > 0$.

4. Toutefois, cette distinction n'est effective que lorsque le nombre de joueurs dépasse 2 ; or Bolton et Ockenfels (2000) se concentrent sur l'analyse de jeux à 2 joueurs (dictateur, ultimatum etc.), ou bien de jeux très compétitifs (Bertrand, Cournot etc.) où la partie « non strictement égoïste » de la fonction d'utilité ne pèse pas sur la décision. On ne peut donc pas clairement identifier si cette différence dans la manière de définir les composantes de la fonction d'utilité, aboutirait à des conclusions différentes. Il aurait été intéressant que Bolton et Ockenfels (2000) traitent le cas d'un jeu du bien public à $n > 2$ joueurs comme le font Fehr et Schmidt (1999).

Dans l'expression décrite dans l'équation (3.2.3), σ_i est égale à :

$$\sigma_i = \begin{cases} x_i/c & \text{si } c > 0 \\ 1/n & \text{si } c = 0 \end{cases}$$

Ainsi, σ_i représente la part relative des gains dans le jeu pour le joueur i , où $c = \sum_{i=1}^n x_i$ est le total des gains monétaires pour l'ensemble des joueurs.

Avec cet exemple, nous pouvons présenter l'ensemble des caractéristiques que les auteurs souhaitent intégrer à la fonction d'utilité d'un agent. Tout d'abord, celle-ci dépend positivement de ses gains monétaires, à travers le premier terme $a_i x_i$. Par ailleurs, la fonction d'utilité intègre également des considérations moins égo-centrées dans la suite de l'expression, en tenant compte de la différence entre, d'un côté, la part relative des gains du joueur i au sein du groupe, σ_i , et, de l'autre côté, la part que chaque joueur obtient lorsque l'ensemble des gains est également réparti entre les joueurs, $\frac{1}{n}$ (avec $n = 2$, ici).

Le deuxième terme de la fonction introduit donc le rapport que chaque agent i entretient vis-à-vis de la norme d'équité. En effet, lorsque $\sigma_i \neq \frac{1}{2}$ (qu'il s'agisse d'une différence en faveur ou en défaveur du joueur i), l'agent i subit le coût psychologique $\frac{b_i}{2} (\sigma_i - \frac{1}{2})^2$, ce qui implique que celui-ci a une préférence pour le fait d'obtenir le gain égalitaire (qui correspond au critère au regard duquel le respect de la norme d'équité est évalué dans ce modèle).

Ici encore, on peut observer la similarité qui existe entre la conceptualisation d'un tel coût psychologique et la douleur qui accompagne l'émergence, d'une part, des émotions morales de colère, de dégoût ou de mépris, lorsqu'une norme morale ou sociale (comme la norme d'équité) est violée à nos dépens et, d'autre part, des émotions comme la culpabilité ou la honte lorsqu'une norme morale ou sociale est violée, à notre avantage.

Cependant, à la différence de Fehr et Schmidt (1999), le coefficient b_i , qui est relatif à l'intensité du coût lié au non respect de la norme, s'applique indifféremment, que cette non conformité soit à l'avantage du joueur i ou à son désavantage. Du point de vue de l'émergence des émotions morales, il n'y a, *a priori*, aucune raison pour que l'impact subi du fait des émotions de condamnation d'autrui, soit de même amplitude que celui qui est issu des émotions morales auto-conscientes à valence négative. Dans cette perspective, une modélisation plus proche de celle réalisée par Fehr et Schmidt (1999) semble plus judicieuse. Toutefois, la formalisation proposée par Bolton et Ockenfels (2000) permet une utilisation plus aisée des outils mathématiques d'optimisation, ce qui peut s'avérer être un choix séduisant dans l'analyse de situations où l'ensemble des stratégies est vaste ou que de nombreuses interactions stratégiques ont lieu entre les agents.

Il convient de noter que les deux coefficients a_i et b_i permettent de caractériser la personnalité, ou le type de l'individu, qui peut se placer entre deux cas extrêmes : strictement « égalitaire », lorsque $a_i = 0$, et strictement égoïste, lorsque $b_i \rightarrow 0$, ce qui permet d'introduire de l'hétérogénéité entre les agents, au sein du modèle⁵.

5. Par ailleurs, en transposant cela au modèle de Fehr et Schmidt (1999), on remarque que ces derniers

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Une autre spécificité de cette formalisation, est que celle-ci pose l'hypothèse d'une information incomplète. En effet, il est supposé que le type de l'agent est une information privée, connue de lui seul. Ce type se caractérise, d'une part, à l'aide de la valeur prise par les paramètres a_i et b_i , et, d'autre part, à travers la définition des deux variables seuils $r_i(c)$ et $s_i(c)$, dont les expressions sont les suivantes :

$$r_i(c) = \underset{\sigma_i}{\operatorname{arg\,max}} U_i(c.\sigma_i, \sigma_i), \quad c > 0$$

$r_i(c)$ désigne la part du total des gains dans le jeu, que le joueur i s'octroierait, s'il était seul à en décider la répartition⁶. Ainsi, d'après la définition même de la fonction d'utilité, (équation (3.2.3)), l'obtention d'une part du total des gains qui dépasse cette valeur seuil impliquerait une désutilité à l'agent, du fait de son aversion à l'iniquité.

Par ailleurs, $s_i(c)$ est la solution de l'équation :

$$U_i(c.s_i, s_i) = U_i(0, \frac{1}{n})$$

Avec $c > 0$ et $s_i \leq \frac{1}{n}$.

Ainsi défini, $s_i(c)$ correspond également à une valeur seuil. En effet, on suppose que lorsque le joueur i obtient, à l'équilibre, une part du total des gains du jeu qui est inférieure à $s_i(c)$, celui-ci préfère obtenir un gain nul, pourvu qu'il en soit de même pour les autres joueurs, plutôt que de se satisfaire d'une telle répartition (ce qui est comparable au seuil de rejet d'une offre dans le jeu de l'ultimatum).

Par conséquent, les individus acceptent d'obtenir, à l'équilibre, une part des gains totaux qui se trouvent entre ces deux seuils (de sorte à ne pas avoir trop, soit pas plus que r , mais à tout de même obtenir suffisamment, soit plus que s). Entre ces deux seuils, on considère que les agents préfèrent plus de gains à moins et donc qu'ils préfèrent se trouver sur le seuil r .

Les auteurs supposent que chaque agent ignore la valeur exacte des seuils $r(c)$ et $s(c)$, caractérisant les autres joueurs. Néanmoins, chacun d'eux connaît la distribution des valeurs prises par ces variables au sein de la population.

En replaçant cet aspect de la modélisation dans la perspective des émotions morales, on peut considérer que ces seuils sont ceux au-delà desquels les émotions qui émergent deviennent trop puissantes (et le coût psychologique trop intense) et conduisent à un arrêt de l'action de recherche du gain matériel le plus élevé possible. Par exemple, dans le cas du jeu de l'ultimatum, une telle représentation implique que chaque individu possède une

posent implicitement l'hypothèse $a_i = 1$. Fehr et Schmidt (1999) semblent ainsi ignorer le cas d'un individu ayant une préférence stricte pour l'équité.

6. Rappel : $x_i = c.\sigma_i$.

valeur « plancher » concernant la répartition des gains proposée par son partenaire, et que si la part qui lui est offerte est inférieure à ce seuil, la colère qui en résulte (nourrie par le sentiment d'injustice) est telle qu'il ignore l'aspect matériel et préfère rejeter l'offre. En ce sens, une telle modélisation est en mesure d'offrir une explication relativement réaliste des comportements de rejet observés expérimentalement dans le jeu de l'ultimatum.

Par ailleurs, l'hypothèse selon laquelle les agents connaissent la distribution des différentes valeurs possibles pour ces seuils, au sein de la population, peut sembler pertinente en ce sens que cette connaissance pourrait refléter une certaine généralisation de ce que les agents ont pu observer, précédemment, au sein de leur environnement. De plus, la communication, mise en place lors d'une négociation par exemple, peut s'envisager comme un moyen d'obtenir plus d'informations sur ces différentes valeurs seuils, voire de permettre une révélation des types des agents. Ce qui pourrait expliquer pourquoi une communication (stratégique) ayant lieu entre les joueurs, avant la participation à un jeu de l'ultimatum, permet de faciliter la coopération entre les partenaires (Zultan, 2012).

3.2.2. Le rôle des intentions et des attentes

Les deux modélisations que nous venons de voir tiennent compte de la sensibilité des agents au fait que l'issue d'un jeu soit en conformité par rapport à une norme (d'équité) établie dans l'économie. Néanmoins, elles ignorent le fait que les agents sont également sensibles aux intentions qui sont sous-jacentes aux actions choisies par ceux avec qui ils interagissent. Blount (1995) montre par exemple que les répondants, dans un jeu de l'ultimatum, ne se comportent pas de la même manière lorsqu'ils sont face à une offre choisie par un autre individu, que lorsqu'elle est décidée par un processus aléatoire (ordinateur). La même offre sera ainsi plus facilement rejetée si elle provient du choix délibéré d'une personne, plutôt que si elle le fruit du hasard. Les résultats obtenus par Falk, Fehr et Fishbacher (2003) ou par Sutter (2007) témoignent également de ce phénomène.

Dans l'expérience de Falk *et al.* (2003), les agents participent à une version discrète du jeu de l'ultimatum dans laquelle l'offreur ne peut choisir qu'entre deux propositions de partage, cette information étant commune aux deux négociateurs. Ainsi, le répondant peut déterminer l'intention de l'offreur, par comparaison entre le choix de partage retenu par ce dernier et l'alternative qui était à sa disposition. Les résultats montrent que la réaction du répondant à une même proposition de l'offreur diffère selon l'offre alternative disponible, ce qui n'est pas compatible avec une explication uniquement fondée sur une préférence pour l'équité.

Dans un protocole similaire, Sutter (2007) montre que l'offre inéquitable (80, 20), toujours dans le cadre du jeu de l'ultimatum, est significativement plus rejetée, lorsque le répondant sait que l'offreur a l'opportunité de proposer l'offre équitable (50, 50). Ces phénomènes de rejet sont le fruit de l'émergence de la colère du répondant, une émotion qui, d'après Ortony, Clore et Collins (1988), survient lorsque l'on est victime d'un préjudice (notamment quand il est perçu comme injuste), que celui-ci soit effectif ou potentiel, et qu'on en attribue la responsabilité, ainsi que l'*intention* à un offenseur identifié.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Néanmoins, ceux qui nous entourent ne nourrissent pas uniquement de mauvaises intentions à notre égard, et il arrive parfois que des personnes décident, de leur plein gré, de réaliser des actions qui nous sont profitables, comme, par exemple, lorsque l'on reçoit un cadeau. Dans ces situations, une autre émotion morale très importante intervient, il s'agit de la gratitude. Comme nous l'avons souligné précédemment, la gratitude est définie comme l'émotion (positive) que l'on ressent lorsqu'une autre personne a effectué ou a tenté d'effectuer un acte, de manière *intentionnelle et volontaire*, entraînant un gain particulièrement précieux (un don d'argent, de l'aide), qui représente un coût pour la personne qui le réalise, sans compensation financière (immédiate) de la part de celui qui le reçoit (McCullough, Kilpatrick, Emmons et Larson, 2001).

Rabin (1993) intègre dans son modèle ces différentes spécificités et donne un rôle central aux intentions des joueurs dans le processus de décision de l'individu. Il suppose, en effet, que l'agent a tendance à être bienveillant avec les personnes qui se sont montrées bienveillantes envers lui, et malveillant avec les personnes qui se sont mal comportées avec lui. Par conséquent, Rabin (1993), dans la lignée des travaux pionniers de Geanakoplos, Pearce et Stacchetti (1989), suppose que l'utilité d'un agent économique, lorsqu'il joue, est affectée par ses croyances à propos des intentions des autres joueurs. Plus précisément, son utilité dépend à la fois de sa stratégie, de ses croyances sur la stratégie de l'autre joueur⁷, et de ses croyances à propos des croyances de l'autre joueur concernant sa propre stratégie.

En reprenant la notation utilisée par l'auteur, on définit $a_i \in S_i$ et $a_j \in S_j$ représentant les stratégies choisies par les joueurs, $b_i \in S_i$ et $b_j \in S_j$ représentant respectivement, les croyances du joueur j sur la stratégie jouée par i , et les croyances du joueur i sur la stratégie choisie par j . Enfin, $c_i \in S_i$ et $c_j \in S_j$ désignent respectivement les croyances du joueur i concernant les croyances que le joueur j a, au sujet de la stratégie jouée par le joueur i , et les croyances du joueur j sur les croyances que le joueur i a, relativement à la stratégie jouée par le joueur j .

Rabin (1993) donne ensuite une définition du paiement équitable, déterminé conditionnellement aux croyances des joueurs sur la stratégie employée par l'autre, qui permet aux agents de déterminer leur croyance sur la bienveillance de leur propre stratégie (à l'encontre de l'autre joueur), mais aussi de celle de l'autre joueur.

Du point de vue du joueur i , ce paiement équitable est égal à :

$$\pi_j^e(b_j) = \frac{\pi_j^h(b_j) + \pi_j^l(b_j)}{2} \quad (3.2.4)$$

Ce gain correspond à la moyenne entre le paiement maximum et le paiement minimum, au sein des allocations pareto-efficaces, que d'après i , j pense atteignables, étant donné que i suppose que l'autre joueur adopte la stratégie b_j . Ainsi, les intentions de i sont bienveillantes (malveillantes) à l'encontre de j si le paiement de j , pour la stratégie choisie par i , est plus grand (petit) que le paiement équitable $\pi_j^e(b_j)$.

7. Rabin (1993) n'envisage que des situations à deux joueurs, néanmoins, Dufwenberg et Kirchsteiger (2004) présentent une extension de ce modèle pour n joueurs.

3.2. L'intégration implicite des émotions morales dans les modèles économiques

De plus, des fonctions de bienveillances, notées $f(a_i, b_j)$ et $f(a_j, b_i)$, définissent respectivement la croyance de l'agent i sur son propre niveau de bienveillance, lorsqu'il joue a_i et qu'il pense que la stratégie de l'autre joueur est b_j et à la croyance de l'agent j sur sa bienveillance, lorsqu'il joue a_j et qu'il pense que la stratégie de l'autre joueur est b_i ⁸. Du point de vue du joueur i , cette fonction est négative s'il donne au joueur j moins que le paiement équitable, positive s'il donne plus et nulle s'il donne exactement le paiement équitable au joueur j .

L'utilité de l'agent est alors donnée par la fonction suivante :

$$U_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + \tilde{f}_j(b_j, c_i)[1 + f_i(a_i, b_j)] \quad (3.2.5)$$

Où $\tilde{f}_j(b_j, c_i)$ correspond à la croyance du joueur i sur la bienveillance que le joueur j pense avoir à propos du joueur i . Par conséquent, si i croit que j considère qu'il ne le traite pas correctement en réalisant sa stratégie b_j ($\tilde{f}_j < 0$), alors i souhaite faire preuve de réciprocité négative en étant également malveillant à l'encontre de j et en choisissant a_i de sorte à ce que f_i soit négatif ou faible. Inversement, si le joueur i pense que j est bienveillant avec lui, il souhaite traiter correctement le joueur j et faire preuve de réciprocité positive.

Chez Rabin (1993), on retrouve à nouveau la présence de la norme, à travers la définition du paiement équitable qui sert de point de référence pour déterminer si les joueurs respectent ou non la norme d'équité. Dans la perspective des émotions morales, l'un des aspects majeurs de cette modélisation, est le parallèle que l'on peut effectuer entre l'interprétation donnée par l'auteur des caractéristiques de sa fonction d'utilité, et les modes de fonctionnement qui ont pu être dégagés, à propos de certaines émotions morales.

Par exemple, si i croit que j ne le traite pas correctement, c'est-à-dire que j choisit délibérément une stratégie conduisant à des gains qui ne respectent pas la norme d'équité, i subit une baisse de satisfaction (puisque $\tilde{f}_j < 0$), pouvant s'apparenter à l'émergence du sentiment de colère. La fonction d'utilité est telle que, étant donné le mauvais comportement de j , i souhaite être également malveillant à son encontre, ce qui le pousse à choisir une valeur de a_i telle que f_i soit négatif ou faible, exactement comme la colère conduit, à travers sa tendance à l'action, à entreprendre de se venger (punitions).

De plus, des recherches ont montré que la vengeance procure un certain plaisir à l'individu qui l'effectue (Rilling et Sanfey, 2011) et ce phénomène se retrouve bien au niveau de la fonction d'utilité, puisque le fait de se venger permet d'atténuer la perte d'utilité due aux mauvaises intentions de l'autre joueur.

8. $f(a_i, b_j) = \frac{[\pi_j(b_j, a_i) - \pi_j^e(b_j)]}{[\pi_j^h(b_j) - \pi_j^{min}(b_j)]}$ où $\pi_j^{min}(b_j)$ est le pire paiement que j peut obtenir lorsqu'il joue b_j .

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Inversement, si le joueur i pense que j est bienveillant avec lui, en lui donnant, intentionnellement, plus que le paiement équitable, i ressent une hausse d'utilité ($\tilde{f}_j > 0$), semblable à celle que l'on peut ressentir lors de l'émergence du sentiment de gratitude. Le joueur i souhaite alors, à son tour, être bienveillant avec le joueur j , de manière similaire à la tendance à l'action de l'émotion de gratitude.

On note ici la pertinence du modèle de Rabin (1993) qui, contrairement aux modèles vus précédemment, intègre une composante de réciprocité positive, et préfigure l'intégration du rôle des émotions positives dans la prise de décision des agents.

De plus, toujours en supposant que le joueur i anticipe des intentions et un comportement bienveillants de la part de son partenaire ($\tilde{f}_j > 0$), on constate que si le joueur i agit de manière malveillante, à l'encontre du joueur j , cela implique une désutilité qui peut être interprétée comme l'émergence de la culpabilité. Ainsi, la présence d'un coût psychologique lié au fait de mal agir à l'encontre d'un partenaire ayant de bonnes intentions à notre égard, peut être capable de dissuader l'adoption d'un tel comportement. Chez Rabin, le sentiment de culpabilité ne peut cependant être envisagé qu'à travers le prisme de la réciprocité.

Par ailleurs, d'après Rabin (1993), la fonction d'utilité définie dans l'équation (3.2.5) permet plus de réalisme qu'une fonction dont la forme serait la suivante :

$$V_i(a_i, b_j, c_i) = \pi_i(a_i, b_j) + \tilde{f}_j(b_j, c_i)f(a_i, b_j)$$

En effet, il explique que lorsqu'une personne n'est pas traitée correctement, c'est-à-dire que l'autre joueur n'a pas de bonnes intentions à son égard, son utilité doit obligatoirement en subir un coût (et donc être inférieure à ses gains monétaires dans le jeu). La vengeance qu'il a l'opportunité d'effectuer en étant malveillant envers l'autre joueur, ne peut donc que partiellement compenser cette perte d'utilité.

Un autre aspect important de la modélisation de Rabin (1993), est l'arbitrage existant, au sein de la fonction d'utilité, entre le bien-être matériel (premier terme de la fonction) et la préférence pour le respect d'une norme, comme celle de l'équité (représentée par le deuxième terme). En particulier, l'auteur suppose que la poursuite du gain matériel peut surpasser l'intérêt pour l'équité et la justice, du moment que l'incitation matérielle est suffisamment forte, ce que l'on retrouve empiriquement, par exemple pour les niveaux de rejet dans le jeu de l'ultimatum, qui diminue proportionnellement, au fur et à mesure que la somme en jeu est importante (Hoffman, McCabe et Smith, 1996 ; Slonim et Roth, 1998 ; Cameron, 1999 ; Munier et Zaharia, 2003).

Cette dualité entre préférence pour le bien-être matériel et préférence pour le fait d'adopter une conduite morale, de sorte à être accepté socialement, est au cœur de ce qui constitue les dilemmes sociaux. En effet, s'il ne s'agissait pas de situations où les différentes préférences des individus rentrent en contradiction les unes avec les autres, elles ne seraient pas qualifiées de dilemmes. De fait, c'est bien de la négligence des préférences sociales ou morales des individus, dans le modèle classique de la théorie des jeux, que provient l'incapacité dont fait preuve cette conception à rendre compte du comportement des agents économiques dans ces situations.

Une telle dualité se retrouve dans la plupart des modèles dits de préférences sociales et on peut, en particulier, souligner à nouveau le travail de Brekke *et al.* (2003), qui centre son analyse sur cette question. Les auteurs reprennent la définition de deux modes de prise de décision proposée par March et Olsen (1995). Le premier se rapproche du fonctionnement classique de l'homo oeconomicus et conclut que l'action individuelle des agents dépend de la réponse aux trois questions suivantes : « Quelles sont les différentes alternatives possibles ? Quelles sont les conséquences de ces alternatives ? Quelle est la valeur, en termes de préférence, de ces conséquences (pour celui qui doit choisir l'action) ? ». Le second mode est quant à lui relatif à l'identité de l'agent et à sa conception des comportements qui sont appropriés à cette identité. Le choix individuel d'une action dépend cette fois-ci de la réponse aux trois questions suivantes : « Quel type de personne suis-je ? Dans quel type de situation suis-je ? Qu'est-ce qu'une personne de mon type ferait dans cette situation ? ». Brekke *et al.* (2003) envisagent alors un procédé hybride, en postulant que les individus veulent d'une part réaliser des actions qui leur permettent de se considérer comme des personnes moralement et socialement responsables, mais sans négliger d'autre part leur désir de bien-être égoïste. Ainsi, au moment de faire leur choix les agents se posent finalement les trois questions suivantes : « Quel type de personne suis-je ? Dans quel type de situation suis-je ? Qu'est-ce qu'une personne comme moi ferait *idéalement* dans ce type de situation ? ». Le choix final résultant d'un arbitrage entre le souhait d'être socialement responsable (en respectant son idéal moral) et la recherche de plaisirs individuels.

Ceci pose la question de savoir comment les être humains réagissent face aux normes. Peut-on acheter la morale, ou la conscience d'un être humain ? Un homme est-il prêt à oublier ce en quoi il croit, ses valeurs, et les normes qu'il a intériorisé au plus profond de lui-même, du moment qu'il y trouve un intérêt matériel égoïste suffisamment important ? Le modèle classique en théorie des jeux néglige complètement le fait que les êtres humains puissent valoriser autre chose que leur bien-être personnel et matériel. Les extensions de ce modèle, en dépit de la prise en compte des préférences sociales des individus, semblent toujours poser un regard cynique sur la nature humaine. Cependant, d'après Orléan (1997), le fait que les individus puissent succomber à l'attrait de gains matériels substantiels pourrait ne représenter qu'une utilisation intelligente de la norme, préconisant de savoir ignorer celle-ci, lorsque la situation à laquelle on fait face l'exige, « autrement dit, il existe une norme supérieure qui autorise, et quelquefois contraint, à violer les règles sociales dans certaines situations exceptionnelles. Ainsi, dans les faits, il existe des marges de choix entre intérêts individuels et respect des normes » Orléan (1997, p. 22). De son point de vue, une déviation, à la marge, vis-à-vis de la norme, dans une situation exceptionnelle, lorsque l'intérêt individuel devient trop prégnant, n'irait par conséquent pas à l'encontre d'une profonde conviction et du respect, de la norme en question, dans des situations moins extrêmes.

Malgré la qualité indéniable de la contribution de Rabin, à travers l'idée, commune aux modèles de la théorie des jeux psychologiques, que les choix des agents dépendent des intentions que l'on attribue aux autres (éventuellement à travers l'influence des émotions morales), certaines faiblesses persistent dans sa modélisation. Lopez-Pérez (2008) explique, en effet, que la définition du paiement équitable de Rabin pose problème. Il prend l'exemple suivant : on se place dans un jeu à deux joueurs et on suppose que

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

$(2,1,0)$ et $(2,2)$ sont les deux seules allocations pareto-optimales disponibles, où la notation (x_1, x_2) désigne respectivement les gains des joueurs 1 et 2. Si la stratégie du second joueur implique le résultat $(2,2)$ - c'est à dire qu'il a, selon la spécification du modèle de Rabin (1993), de mauvaises intentions vis-à-vis du premier joueur - alors le joueur 1 pourrait avoir envie de se venger, s'il en a la possibilité. Dans les faits, ce type de comportement semble relativement irréaliste. Ainsi, à l'instar des modèles d'aversion à l'iniquité qui manquent de généralité du fait qu'ils ne prennent pas en considération les intentions des joueurs, « les modèles d'intention sont incomplets car ils ne prennent pas suffisamment en compte le comportement face à la distribution des gains » (Garapin et Ruffieux, 2009, p. 21).

Dans la continuité du travail de Rabin (1993), Falk et Fischbacher (2006) proposent une extension de son modèle qui s'applique aux jeux ayant une forme extensive. Celle-ci suppose alors que l'utilité de l'agent dépend de ses gains matériels dans le jeu et de son évaluation des intentions de l'autre joueur. Toutefois, là où, dans le modèle de Rabin, les intentions des joueurs sont déterminées sur la seule base des conséquences matérielles qu'elles impliquent à l'issue du jeu, dans le modèle de Falk et Fischbacher (2006), ces intentions sont jugées bienveillantes (malveillantes) à partir de *deux* critères : (1) le critère dit « outcome » qui dépend de la répartition des gains entre les joueurs au nœud considéré. Un joueur estime que les intentions d'un partenaire sont bienveillantes (malveillantes), selon ce critère, lorsque il obtient des gains espérés plus (moins) élevés que l'autre joueur. En complément de cette première indication, (2) le critère dit « d'intentionnalité » permet de déterminer si la stratégie de l'autre agent provient d'un choix totalement délibéré, ou si celui-ci était plus ou moins contraint dans son choix (pouvait-il faire autrement?).

Cette modélisation partage ainsi, avec d'autres (Charness et Haruvy, 2002 ; Charness et Rabin, 2002 ; Erlei, 2008), la caractéristique de constituer une *hybridation* (Garapin et Ruffieux, 2009) entre les modèles à la Fehr et Schmidt (1999) et les modèles d'intention. De manière générale, les modélisations *hybrides* ont tendance à s'avérer plus performantes, vis-à-vis de la confrontation avec des données empiriques, que les modèles qui ne supposent qu'un seul type de préférence. Ces formalisations partagent néanmoins, avec les autres modèles appartenant à la théorie des jeux psychologiques, le défaut d'être relativement sophistiquées au niveau conceptuel.

3.2.3. Le rôle des émotions dans la dynamique des liens sociaux

Les travaux de formalisation que nous avons revus jusqu'ici, préfigurent le rôle statique de l'influence des émotions dans la prise de décision des individus. Que cette influence soit mise en évidence à travers une réinterprétation des formes fonctionnelles étudiées, ou qu'elle ait été envisagée consciemment par les auteurs, bien que de manière relativement vague et imprécise, à un stade précoce de l'intégration formelle des sentiments, celle-ci permet de décrire comment l'anticipation des conséquences liées à l'émergence de certaines émotions (sous formes de coûts ou de gains psychologiques) modère les préférences matérielles égo-centrée des individus, et représente une force orientant le comportement des individus vers ce qui est socialement adapté, pour une période donnée.

3.2. L'intégration implicite des émotions morales dans les modèles économiques

Néanmoins, on trouve également, au sein de la littérature en économie, des travaux dont l'intérêt se porte sur l'analyse théorique de l'influence dynamique des émotions sur les comportements individuels, dans les situations de dilemme social.

En particulier, van Dijk et van Winden (1997) propose un travail précurseur décrivant la façon dont les émotions ressenties par des agents participant à un jeu du bien public, conditionnent l'évolution du lien social existant entre eux, lorsqu'ils jouent de manière répétée. En effet, les auteurs indiquent que, parmi les dimensions utilisées par les sociologues et les psychologues sociaux, pour caractériser un lien social, au sein desquelles on retrouve l'intensité émotionnelle, la mise en œuvre de services réciproques ou la confiance mutuelle, la composante affective est celle qui est considérée comme la plus déterminante.

Fort de cette conviction, van Dijk et van Winden (1997) conceptualisent le lien social existant entre deux individus, ainsi que la force qui le caractérise, sous la forme des sentiments ressentis par chacun des agents à l'encontre de l'autre, étant donné que ces sentiments définissent la mesure dans laquelle chaque agent se soucie du bien-être de son partenaire.

Soit deux agents i et j participant à un jeu du bien public, dont les dotations initiales sont respectivement égales à y_i et y_j . Les auteurs supposent alors que le niveau de satisfaction obtenu par chaque agent $i \in I$, $I = \{i, j\}$, dépend d'une part de la consommation de ses propres gains matériels dans le jeu, noté U_i , et d'autre part de la satisfaction qu'il dérive de la consommation que fait son partenaire des gains matériels dont ce dernier bénéficie, noté U_j .

Plus précisément, pour tout $i \in I$, on a $U_i = U_i(x_i, g)$, où x_i désigne la consommation en bien privé de l'agent i , et où g représente la consommation en bien public qui est réalisée par chacun des deux agents, et U_i est une fonction croissante en chacun de ses arguments (les préférences des individus étant supposées convexes et le prix des deux biens étant égal à 1). De plus, on note g_i la contribution de chaque agent $i \in I$ au bien public, de sorte à ce que $g_i + g_j = g$ et que $y_i = x_i + g_i$.

Par ailleurs, le lien social entre les deux joueurs i et j est caractérisé par la paire $(\alpha_{ij}, \alpha_{ji})$ et permet de définir le poids accordé par chaque agent i à la satisfaction tirée par j de la consommation matérielle que celui-ci réalise. Il convient également de noter que van Dijk et van Winden (1997) supposent que le domaine de définition de α_{ij} se limite à l'intervalle $(-1, 1)$, ce qui implique qu'un individu ne peut pas accorder plus d'importance, que ce soit positivement ou négativement, au bien-être matériel de son partenaire, qu'au sien.

La fonction d'utilité de chaque agent $i \in I$, $i \neq j$, est alors définie comme :

$$V_i = U_i U_j^{\alpha_{ij}} \quad (3.2.6)$$

La maximisation de la fonction (3.2.6) conduit à définir, pour chaque agent $i \in I$, une fonction de meilleure réponse permettant de déterminer sa contribution optimale au bien public, en fonction de sa propre dotation initiale, de celle de son partenaire, de la contri-

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

bution de ce dernier, et des sentiments α_{ij} que i ressent à l'égard de celui-ci. En supposant une solution intérieure, la confrontation des deux fonctions de meilleure réponse permet d'obtenir un unique équilibre tel que $g_i = g_i(y_i, y_j, \alpha_{ij}, \alpha_{ji})$, pour tout $i \in I, i \neq j$.

Les résultats de statique comparative montrent que la contribution des agents varie positivement à mesure que leur propres sentiments à l'égard du partenaire s'améliorent, ou bien que leur dotation initiale est importante, tandis que celle-ci a tendance à diminuer lorsque l'autre individu ressent des sentiments plus positifs, ou que ce dernier bénéficie de ressources plus importantes.

Par ailleurs, lorsque l'on suppose que les sentiments ressentis par les agents, l'un envers l'autre, ont une intensité nulle ($\alpha_{ij} = \alpha_{ji} = 0$), on retrouve le modèle classique en théorie des jeux ($V_i = U_i(x_i, g)$) et l'équilibre de Nash obtenu correspond à ce qui est généralement observé expérimentalement, à savoir un niveau de contribution positif de la part des agents, mais qui reste en dessous de ce qui est souhaitable socialement. Face à ce « benchmark », les auteurs indiquent que dans le cas où les individus ont les mêmes préférences ($V_i = V_j$), et que le lien social existant entre les joueurs a une intensité maximale, ($\alpha_{ij} = \alpha_{ji} \rightarrow 1$), le niveau de contribution qu'ils réalisent correspond à celui qui maximise une fonction de bien-être social. A l'inverse, si le lien social entre eux est d'intensité minimale ($\alpha_{ij} = \alpha_{ji} \rightarrow -1$), la contribution des agents l'est également, et celle-ci peut même être nulle lorsque les joueurs disposent d'un même niveau de ressources.

Ainsi, une des implications majeures du modèle de van Dijk et van Winden (1997), est que la spécificité du lien social unissant deux agents, qui dépend de la nature des sentiments qu'ils ressentent l'un envers l'autre, est déterminante dans leur capacité à mettre en œuvre une action coopérative. De plus, d'après leurs résultats, il semble que le fait de resserrer le lien social existant entre les individus permet d'orienter leurs comportements vers une maximisation du bien être social, ce qui est en accord avec les éléments expérimentaux revus lors du chapitre 2.

Néanmoins, la contribution la plus cruciale des auteurs se rapporte au fait que ceux-ci conceptualisent l'évolution du lien social de manière dynamique, en supposant que les sentiments ressentis par les individus, les uns envers les autres, se développent au cours du temps, avec la répétition prolongée des interactions interpersonnelles.

Plus précisément, l'hypothèse est ici que le comportement contributif mis en œuvre par un individu, de part le fait qu'il représente une externalité positive pour autrui, est susceptible de faire évoluer les sentiments qui sont ressentis à son égard. Toutefois, cette évolution sera différente selon la manière dont le comportement de contribution d'un agent se compare avec celui de son partenaire. En effet, les auteurs supposent que les sentiments suscités chez l'agent i sont à valence positive (négative) lorsque le niveau de contribution de l'agent j est supérieur (inférieur) au sien, en tenant compte d'un certain degré de tolérance. Ils indiquent également que cette comparaison des contributions se rapporte à une notion d'équité ou de justice. Formellement, la nature du sentiment suscité chez l'agent i est définie par la relation suivante :

$$G_{ij} \equiv g_j - \varepsilon_i g_i \tag{3.2.7}$$

Où $0 \leq \varepsilon_i \leq 1$ désigne le degré de tolérance de l'individu i , le produit $\varepsilon_i g_i$ correspondant alors à l'attente du joueur i vis-à-vis du comportement contributif d'autrui.

Bien que van Dijk et van Winden (1997) ne désignent pas spécifiquement les émotions qui sont à l'œuvre dans leur modèle, on peut par exemple, s'agissant du cas où un agent observe l'adoption d'un comportement injuste de la part de son partenaire, reconnaître le contexte d'émergence du sentiment de colère. De la même façon, il est aisé de relier l'émergence de l'émotion morale de gratitude chez un agent, au fait que celui-ci soit témoin d'une contribution supérieure à ses attentes, de la part d'un partenaire. Enfin, le fait de supposer que l'émergence de ces émotions est conditionnée par un tel processus de comparaison des contributions, se rapproche sans doute du fait de supposer l'existence d'une norme régissant les comportements de contribution, et qui veut que les agents doivent se comporter sensiblement (du fait de la présence des divers degrés de tolérance) de la même façon. Ce modèle est donc une nouvelle illustration de la manière dont le rôle des émotions et des normes morales, que nous avons souligné lors des chapitres précédents, semble parfaitement prendre part aux mécanismes psychologiques ayant été judicieusement perçus par les économistes, avant qu'ils soient intégrés à leur analyse théorique.

Pour tout agent $i \in I$, $i \neq j$, l'équation différentielle suivante décrit le développement des liens sociaux au cours du temps :

$$\frac{d\alpha_{ij}}{dt} = f(G_{ij}, \alpha_{ij}) \quad (3.2.8)$$

Où f est continue, deux fois différentiable et est une fonction positive de G_{ij} , que les auteurs nomment « l'impulsion ». Cette dénomination est pertinente dans le sens où elle illustre bien la force motrice, parfois inconsciente, des émotions, qui conduit à mieux percevoir la teneur d'une interaction et pousse souvent à agir en conséquence, selon certaines tendances à l'action. Par ailleurs, les auteurs envisagent qu'il peut y avoir une certaine diminution de l'intensité des liens sociaux de période en période, due à une dégradation latente de la puissance des émotions dans le temps. Malgré le bien-fondé de cette hypothèse, nous éludons cet aspect du modèle, afin d'alléger notre présentation.

La fonction f est définie de telle manière à ce que le sentiment α_{ij} reste toujours dans son domaine de définition $(-1, 1)$. De plus, dans la cas où il n'y a pas de dégradation des sentiments, il est supposé que lorsque $G_{ij} = 0$, alors $f(G_{ij}, \alpha_{ij}) = 0$ et l'intensité du lien social existant pour i , vis-à-vis de j reste inchangé, d'une période à l'autre.

Afin de décrire complètement le fonctionnement dynamique du modèle, régissant le développement des liens sociaux, ainsi que l'évolution des contributions que cela implique, à l'équilibre de chaque période, il convient de souligner une hypothèse posée par van Dijk et van Winden (1997) à propos des agents. En effet, il est supposé dans ce modèle que les individus sont myopes, au sens où ils ne prennent pas en compte l'impact potentiel

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

des niveaux de contribution sur l'évolution de leurs propres sentiments, ainsi que de ceux ressentis par les autres joueurs, à leur égard. La structure de décision peut alors être modélisée comme une séquence de décisions de contribution qui ne sont reliées entre elles que par le développement des liens sociaux, au cours du temps, ce qui permet de déterminer la séquence des équilibres de Nash qui sont joués à chacune des périodes considérées, en fonction de la paramétrisation initiale et de l'histoire qui en découle.

La formalisation novatrice de van Dijk et van Winden (1997) est particulièrement intéressante en ce sens qu'elle examine le rôle dynamique des émotions dans l'évolution des relations entre agents économiques, lorsque ceux-ci sont amenés à interagir de manière répétée. En effet, les auteurs considèrent que le développement des sentiments qui a lieu entre les individus, sur la base de leurs comportements de contribution, dans un jeu du bien public, régit l'évolution de l'intensité du lien social existant entre eux et que la mise en œuvre de comportement prosociaux (asociaux), lors d'une période donnée, est en mesure de renforcer (dégrader) ce lien, ce qui influence le comportement stratégique des agents, à la période suivante. Le fait que les agents ne soient pas conscients de la portée de leurs comportements, en ce qui concerne l'évolution des sentiments interpersonnels, au delà de la période qui est considérée, au moment de la prise de décision, nous semble relativement réaliste. Il semble en effet plutôt sensé de considérer que les processus émotionnels puissent être difficiles à saisir et à anticiper sur la durée (soit d'une période à l'autre).

Néanmoins, les auteurs restent évasifs sur la nature des émotions qui interviennent dans leur modèle et, par ailleurs, leur représentation implique qu'à chaque période, les agents prennent leur décision sans se soucier de la portée morale de l'action qu'il mette en œuvre, ni même de l'appropriabilité sociale de leur comportement.

Certes, à travers les fonctions de meilleure réponse, chaque agent adapte sa stratégie pour qu'elle soit optimale, étant donnée celle qui est choisie par le partenaire. De plus, à travers la prise en compte des sentiments ressentis par les agents, à propos de leurs partenaires, cette formalisation tient compte du fait que les êtres humains font preuve de sensibilité quand ils interagissent avec autrui et que les sentiments jouent dans leur considération d'un partenaire.

Pour autant, dans ce modèle, aucun agent ne cherche à porter un jugement moral sur le choix (et les intentions) d'un partenaire. De même, aucun agent ne tient compte de la manière dont son comportement se compare aux attentes d'autrui, à son jugement (dés)approbateur, ni même à un quelconque impératif moral qui serait intériorisé par les individus. Or, lors d'une interaction sociale, nous soutenons que les individus ont conscience de la portée morale et sociale de leurs choix et, par conséquent, une représentation des individus qui néglige cette dimension est sans doute relativement limitée. Aussi, nous pensons que l'approche développée par van Dijk et van Winden (1997) mérite d'être envisagée sous un angle qui tient également compte de l'influence des normes et des multiples émotions morales participant à la prise de décision des agents économiques, dans les dilemmes sociaux.

Avec ce dernier modèle, on constate à nouveau que le rôle des émotions a parfois nourri la création de modèles originaux, étant en mesure de représenter de manière plus réaliste le comportement des êtres humains. Cependant, cette inspiration a pu sans doute être

limitée par le manque de recherches et de résultats empiriques concernant la spécificité des émotions, ne permettant ainsi qu'une approche relativement intuitive des mécanismes psychologiques sous-jacents à la prise de décision économique.

Nous nous tournons maintenant vers des travaux de modélisation qui ont pu bénéficier des avancées réalisées dans les domaines de l'économie expérimentale et de la psychologie sociale (voire des neurosciences) pour affiner leur représentation des agents économiques en intégrant sciemment et spécifiquement le rôle de certaines émotions morales.

3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique

Au sein des modèles que nous avons pu voir jusqu'ici, rares sont les références explicites et spécifiques ayant été faites au rôle particulier d'une ou plusieurs émotions morales. Il est fort possible que certains des auteurs que nous avons mentionnés aient correctement identifié les mécanismes psychologiques à l'œuvre lors du processus de décision, sans pour autant avoir été en mesure, ou même éventuellement avoir souhaité, faire référence aux affects ou aux émotions. Cependant, du fait de la similitude entre les mécanismes présentés dans les différents modèles étudiés et les mécanismes inhérents au rôle des émotions morales (vu dans les chapitres précédents), ces formalisations représentent clairement une première étape vers l'intégration de ces émotions dans la conception du processus de décision, par les économistes.

Plus récemment, certains travaux de formalisation ont franchi le pas, en identifiant explicitement le rôle spécifique joué par des émotions morales particulières. L'avantage que ces modèles présentent, par rapport à ceux qui n'intègrent qu'implicitement le rôle des sentiments moraux, provient en grande partie du fait qu'à travers l'identification précise des émotions en actions, les auteurs peuvent affiner leur représentation des mécanismes à l'œuvre lors du processus de décision, sur la base des travaux expérimentaux (en économie et en psychologie) qui sont relatifs à ces émotions particulières. Ainsi, l'intégration explicite du rôle des émotions morales dans les modèles de représentation du processus de décision des agents constitue un raffinement offrant potentiellement un niveau supérieur de réalisme.

3.3.1. Les modèles centrés sur le rôle d'une émotion particulière

Au sein de la littérature en économie, on constate que les travaux de modélisation qui ont choisi de s'intéresser à la thématique des émotions se sont souvent centrés sur l'analyse spécifique d'une émotion particulière. Parmi les émotions traitées, la honte et la culpabilité sont celles qui ont sans doute le plus réussi à captiver l'intérêt des modélisateurs. Toutefois, d'autres chercheurs, plus rares, ont aussi étudié le rôle d'émotions à valence positive comme la gratitude. Par ailleurs, d'autres travaux, plus ambitieux, ont souhaité relever le défi d'intégrer simultanément le rôle de plusieurs émotions, en tentant d'identifier et d'intégrer la subtilité de leurs disparités.

3.3.1.1. Un exemple de modèle d'aversion à la culpabilité

Parmi les modèles d'aversion à la culpabilité, on retrouve celui de Battigalli et Dufwenberg (2007). Cette formalisation se base en partie sur l'expérimentation réalisée par Dufwenberg et Gneezy (2000). Celle-ci cherche à étudier le comportement des individus lorsqu'ils sont face au choix suivant : un agent (le joueur 1) trouve un porte-monnaie par terre dans la rue, celui-ci peut alors le garder, auquel cas il récupère la certaine somme d'argent présente dans le porte-monnaie, ou alors le ramener au poste de police qui assure la liaison et qui offre ensuite l'opportunité au propriétaire (le joueur 2) de prendre contact avec lui et de lui donner une somme d'argent afin de le dédommager. Si le joueur 1 rapporte le porte-monnaie, le joueur 2 doit alors faire le choix du montant qu'il souhaite lui renvoyer. Les auteurs montrent que le montant renvoyé par le joueur 2, est corrélé avec la croyance que le joueur 2 possède à propos des attentes du joueur 1 vis-à-vis de la somme à renvoyer. Battigalli et Dufwenberg (2007) se servent de ce résultat pour définir la manière dont la culpabilité émerge et intervient dans leur modèle.

On considère alors un jeu à n joueurs. Nous nous plaçons du point de vu du joueur i . Dans le modèle, la source de la culpabilité ressentie par le joueur i , vis-à-vis de chacun de ses partenaires, est liée aux attentes de chaque joueur $j \neq i$. En effet, les auteurs supposent qu'étant donné la stratégie s_j choisie par un joueur j , le joueur i détermine, pour chaque nœud terminal du jeu atteignable, une croyance D_j sur le niveau de « déception » ressenti par ce partenaire. Ce niveau de déception correspond à la différence entre l'espérance de gain du joueur j , par rapport à chaque nœud terminal possible du jeu lorsque celui-ci choisit s_j , et le gain monétaire obtenu pour le nœud terminal considéré. Si cette différence est positive, le joueur i anticipe un niveau de déception positif chez j , si cette différence est négative ou nulle, le niveau de déception est égal à zéro. Plus simplement, le joueur i anticipe un niveau de déception positif de l'autre joueur lorsqu'il croit que celui-ci obtient moins que ce qu'il pouvait espérer. A l'inverse, il anticipe un niveau de déception nul s'il croit que le joueur j obtient une somme supérieure ou égale à son gain espéré.

A partir de ce niveau de déception global D_j , le joueur i est capable de déterminer le niveau de déception de j dont il est lui même responsable du fait de la stratégie qu'il adopte, noté G_{ij} . En utilisant une notation simplifiée, on peut écrire : $G_{ij} = D_j - \min_{s_i} D_j$. Ainsi, pour une stratégie s_i donnée, le niveau de déception que i peut s'attribuer correspond à la différence entre le niveau global de déception qu'il anticipe pour j , lorsqu'il choisit la stratégie s_i , et le niveau global de déception minimal attribué à j , en considérant l'ensemble des autres stratégies possibles pour i .

Enfin, avec une notation légèrement simplifiée, la fonction d'utilité de l'agent i , conditionnellement au nœud terminal z , est définie de la manière suivante :

$$u_i^{SG}(z) = m_i(z) - \sum_{j \neq i} \theta_{ij} G_{ij}(z) \quad (3.3.1)$$

Où $m_i(z)$ est le gain monétaire du joueur i au nœud terminal z et θ_{ij} est un paramètre représentant la sensibilité du joueur i au sentiment de culpabilité. Cette hypothèse n'est pas spécifiquement commentée par les auteurs, mais l'indice ij du paramètre θ_{ij} semble

3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique

signifier que l'agent serait capable d'avoir une sensibilité différente à la culpabilité, selon l'agent auquel il fait face, ce qui semble relativement naturel (on ne ressent pas les émotions avec la même intensité lorsque l'on est face à un inconnu que lorsque l'on est face à un proche), bien que cela soit absent dans les modélisations précédemment présentées.

En résumé, dans le modèle de Battigalli et Dufwenberg (2007) l'accent est mis sur les *attentes* des autres individus, dont le rôle dans la prise de décision économique a été notamment mis en évidence expérimentalement par Dana, Cain et Dawes (2006). De plus, les auteurs supposent que les agents cherchent à déterminer quelles sont les attentes d'autrui, vis-à-vis du *résultat* du jeu, et que ceux-ci souffrent de ne pas se conformer à ces dernières, en raison de l'émergence du sentiment de culpabilité.

Cependant, il convient de se demander si les attentes particulières des autres peuvent avoir un caractère normatif. Peut-on considérer que l'on va ressentir de la culpabilité si on ne se conforme pas aux attentes d'un autre individu, alors même que l'on juge ces attentes extravagantes et injustes ? De ce point de vue, cette modélisation semble souffrir d'un manque de réalisme et il serait sans doute plus pertinent de considérer que les agents tentent d'abord de déterminer une norme de comportement pouvant s'appliquer à l'ensemble des individus avec lesquels ils interagissent, dans une situation particulière, et qu'éventuellement, ceux-ci se sentent ensuite coupables lorsqu'ils ne se conforment pas à cette norme.

En outre, une expérimentation de Vanberg C. (2008) remet en question le rôle des attentes dans le processus de décision des agents. L'auteur s'intéresse, en effet, aux raisons pour lesquelles les individus respectent les promesses qu'ils font, avant de commencer un jeu économique, et met en place une expérimentation permettant de discriminer entre deux théories explicatives de ce phénomène. Selon Vanberg C. (2008), la première théorie suppose que les agents ont une préférence pour le fait de respecter leur parole, ce qui implique, plus précisément, que le fait de promettre induit un engagement émotionnel chez l'agent le conduisant ensuite à mener à bien son contrat moral. L'autre théorie explicative qui est proposée se réfère aux attentes et conclut que les agents respectent leur parole en raison d'une aversion au fait de décevoir les attentes (matérielles) développées par ceux à qui ils ont fait une promesse. Cette deuxième théorie implique alors que ce n'est pas tant le fait même qu'un agent promette quelque chose qui pèse ensuite sur sa décision, mais plutôt le fait que cela induise un changement au niveau des attentes de ceux envers qui la promesse a été effectuée.

L'expérimentation proposée par Vanberg C. (2008) réunit des participants, par paire, pour qu'ils jouent à un mini-jeu du dictateur. Ce jeu se déroule en deux étapes. A la première étape, le rôle de chacun des deux joueurs est déterminé au hasard et chaque joueur possède une chance sur deux d'être le dictateur. Ensuite, c'est le dictateur qui joue en premier et celui-ci a deux choix stratégiques possibles. Il peut choisir l'action « roll » (« action de faire rouler le dé »), auquel cas il reçoit de manière certaine un gain de 10 u.e. (unité expérimentale), tandis que l'autre joueur tire un dé et reçoit 12 u.e. avec la probabilité $\frac{5}{6}$ et 0 avec la probabilité $\frac{1}{6}$. L'autre action qui peut être choisie par le dictateur est « don't roll » (« ne pas faire rouler le dé »), auquel cas le dictateur reçoit 14 u.e., quand l'autre joueur obtient un gain nul. De plus, lors d'une étape préalable au commencement du jeu, les individus constituant chaque paire de joueurs ont l'opportunité de communiquer en-

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

semble. Cette opportunité aboutit le plus souvent à une promesse réciproque impliquant, pour chaque joueur, que s'il se retrouve dans la position du dictateur il choisira l'action « roll ».

Afin de déterminer quelle théorie est la plus pertinente pour expliquer la raison qui pousse les agents à respecter une promesse, Vanberg C. (2008) partage son échantillon en deux groupes. S'agissant des participants du premier groupe, le déroulement de l'expérimentation s'effectue comme expliqué ci-dessus. Pour les autres, l'appariement des agents est rompu entre l'étape de communication et la participation au jeu. Ainsi, les dictateurs ne jouent pas avec l'individu avec lequel ils ont communiqué et à qui ils ont potentiellement fait une promesse. De plus, les dictateurs sont les seuls à savoir que les paires ont changées et ceux-ci ont également la possibilité de consulter la discussion qui a eu lieu entre leur nouveau partenaire et le dictateur avec lequel ce dernier était censé jouer. Par conséquent, chaque dictateur de cette condition est en mesure d'observer si une promesse a été faite à celui avec qui il joue dorénavant. Vanberg C. (2008) suppose alors que le choix de l'action « roll », par les dictateurs de cette seconde condition, s'explique plutôt parce que la promesse qui est faite à un partenaire entraîne un changement au niveau des attentes de ce dernier, dans la mesure où ce n'est pas le dictateur lui-même qui est responsable de la promesse en question.

Les résultats montrent que lorsqu'un dictateur effectue lui-même la promesse de jouer « roll », cela modifie significativement son comportement au moment du jeu, de sorte à ce que celui-ci honore plus souvent sa promesse. A l'inverse, pour les individus du second groupe, bien que la promesse faite par un autre joueur mène à un changement significatif des croyances du dictateur vis-à-vis des attentes de son (nouveau) partenaire, cela n'a pas d'influence significative sur son choix stratégique au moment du jeu, et cela même si le dictateur avait également fait une promesse lors de la phase de communication.

Ainsi, cette expérience remet grandement en question le rôle des attentes dans l'émergence du sentiment de culpabilité, comme dans la mise en œuvre de son influence décisionnelle. Sur la base des résultats de Vanberg C. (2008), il semble que l'émergence du sentiment de culpabilité ne soit pas tant liée à l'adoption d'un comportement menant à un résultat qui déçoit les attentes des autres, comme c'est le cas dans le modèle de Battigali et Dufwenberg (2007), mais plutôt que celle-ci soit liée au fait d'adopter un comportement incohérent avec un impératif contractuel ou moral. Par conséquent, il semble que ce soit l'anticipation d'obligations morales, comme les normes, qui influence les comportements, à partir de l'émergence de la culpabilité.

Dans cette perspective, l'approche développée par Battigali et Dufwenberg (2007) pourrait être mieux adaptée à une représentation du rôle de la honte. En effet, Tadelis (2008) explique que la honte est définie comme une émotion qui modifie les comportements d'un individu à travers l'anticipation qu'il fait de ce que les autres vont penser de lui. Le mécanisme incitatif, permettant une modification du comportement de l'agent, trouve son origine dans son aversion au fait d'être perçu comme un individu possédant un mauvais trait de personnalité ou se comportant d'une manière inadaptée.

En s'intéressant à la manière dont l'agent est amené à prendre en compte ce que les autres pensent de ses actions, voire de ce qu'ils peuvent en déduire, à propos de sa personnalité, la modélisation de Battigali et Dufwenberg (2007) se rapproche donc d'un questionnement relatif au contexte d'émergence de la honte. Dans la conception de Tadelis (2008), à la différence de l'émotion de culpabilité, qui agit comme une motivation *intrinsèque* dans la modulation du choix de l'agent, liée aux propres croyances de l'individu sur la manière dont il doit se comporter (et sur les normes qu'il doit respecter), l'émotion de honte est considérée comme une source de motivation *extrinsèque* dans la régulation des comportements de l'agent, qui est en rapport avec le niveau d'exposition publique de ses actions, et qui dénote une sensibilité de l'individu à la perception qu'on les autres concernant sa conduite.

3.3.1.2. Un exemple de modèle d'aversion à la honte

Tadelis (2008) propose un modèle de théorie des jeux psychologiques s'inscrivant dans le contexte d'un jeu de confiance. Dans ce jeu, le joueur 1 prend la première décision et doit choisir s'il souhaite faire confiance (stratégie notée T) ou ne pas faire confiance (stratégie notée N) au joueur 2. Lorsque le joueur 1 choisit de ne pas faire confiance, le jeu s'arrête et chaque joueur obtient un gain v . Si le joueur 1 décide de faire confiance à son partenaire, le joueur 2 a ensuite le choix entre deux stratégies : coopérer (C) ou dévier (D). Lorsque le joueur 2 dévie, celui-ci obtient le gain maximal possible dans le jeu, noté d , tandis que le gain du joueur 1 est nul. Le jeu utilisé par Tadelis (2008) présente une subtilité, par rapport à un jeu de confiance traditionnel, qui est que la coopération ne garantit pas les gains du joueur 1. En effet, si le joueur 2 décide de coopérer, la nature intervient et la réussite de la coopération est alors incertaine. Plus précisément, avec la probabilité $p \in [0, 1]$, le joueur 1 obtient un gain $\frac{c}{p}$, alors qu'avec la probabilité $(1 - p)$, le gain du joueur 1 est nul, de sorte à ce que le gain espéré du joueur 1 soit c . Le joueur 2 obtient quant à lui le gain c quelle que soit l'action de la nature. S'agissant du joueur 1, l'action de faire confiance, lorsque cela est suivi de la coopération, est un profil de stratégie paréto-dominant, par rapport au choix de ne pas faire confiance, puisqu'il obtient alors un gain espéré $c > v$. Toutefois, le choix de coopérer est moins attractif que la stratégie de déviation pour le joueur 2, étant donné que $c < d$. Ainsi, si on suppose que les deux joueurs sont rationnels, égoïstes et qu'ils anticipent correctement l'action de leur partenaire, le joueur 1, étant conscient que le joueur 2 choisit de dévier lorsqu'on lui fait confiance, préfère ne pas faire confiance afin d'obtenir le gain $v > 0$.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Pour plus de clarté, l'arbre du jeu est représenté dans la figure 3.3.1 :

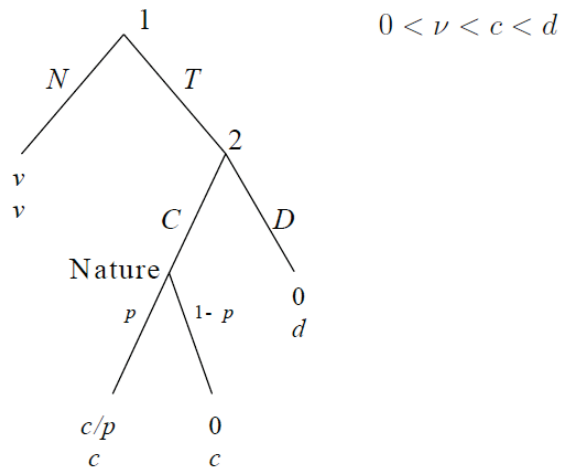


FIGURE 3.3.1.: Arbre du jeu de confiance dans le modèle de Tadelis (2008)

Du fait de l'intervention de la nature, même lorsque le joueur 2 agit de manière coopérative, il y a une chance pour que le joueur 1 obtienne un paiement nul, qui est équivalent à ce qu'il reçoit lorsque le joueur 2 dévie. De fait, lorsque le joueur 1 ne peut observer l'action de son partenaire, celui-ci n'est pas capable de déterminer si le fait d'obtenir un gain nul est lié au comportement stratégique du joueur 2 ou seulement au hasard. Par conséquent, il n'a aucune possibilité d'inférence en ce qui concerne la bonne foi de son partenaire. Cet aspect est central dans la modélisation de Tadelis (2008), puisque l'auteur s'intéresse au (seul) sentiment de honte et suppose que celui-ci est en lien avec le niveau d'exposition des actions choisies par les joueurs.

Le joueur 1 est supposé neutre au risque et égoïste. Ainsi, lorsque celui-ci reçoit un paiement espéré de m_1 , son utilité est égale à :

$$u_1 = m_1 \tag{3.3.2}$$

3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique

On admet que le joueur 2 est également neutre au risque, cependant, celui-ci est également caractérisé par la présence de préférences sociales. Formellement, l'auteur suppose qu'en plus de ses gains matériels, la fonction d'utilité de l'agent 2 prend en considération ses croyances sur les croyances du joueur 1 à propos de la stratégie employée par le joueur 2. En notant $E[\sigma] \in [0, 1]$, la croyance *ex post* du joueur 1 sur la probabilité que le joueur 2 ait choisi de coopérer, la fonction d'utilité du joueur 2 s'écrit comme suit :

$$u_2 = m_2 - (1 - E[\sigma])s \quad (3.3.3)$$

Où m_2 désigne le gain matériel du joueur 2 et où on note s le niveau d'aversion à la honte de ce même joueur.

Comme Tadelis (2008) l'indique, cette spécification implique que le joueur 2 n'est pas averse au fait de dévier, mais ressent plutôt une désutilité lorsque le joueur 1 attribue une probabilité positive au fait qu'il dévie. De même, cela revient à supposer que le joueur 2 ne se soucie pas du sort de son partenaire mais seulement de la manière dont celui-ci le perçoit, ainsi que, par extension, de l'approbation morale qui peut être accordée, en conséquence, à sa conduite.

Par ailleurs, le niveau d'exposition $e \in (0, 1)$ du comportement du joueur 2 conditionne la croyance *ex post* $E[\sigma]$ du joueur 1. En particulier, dans le cas où le joueur 1 reçoit un gain nul, le niveau maximal d'exposition $e = 1$ garantit que celui-ci puisse observer s'il s'agit d'un choix volontaire de déviation du joueur 2 ou si cela résulte de l'action aléatoire de la nature. A l'inverse, le niveau minimal d'exposition $e = 0$ n'offre aucune information à la suite de l'obtention d'un gain nul.

Lors de la résolution du modèle, étant donné la structure de paiement du jeu et le niveau d'exposition du comportement du joueur 2, celui-ci arbitre entre le gain matériel marginal $d - c$ obtenu en déviant et le coût émotionnel consécutif de la déviation. Ceci conduit alors à la définition d'une valeur seuil de l'intensité s , avec laquelle le joueur 2 ressent l'émotion de honte, permettant de déterminer le choix stratégique que celui-ci effectue, lorsque son partenaire lui fait confiance. Ainsi, chaque joueur 2 caractérisé par un intensité d'aversion à la honte inférieure à la valeur seuil choisit D , tandis qu'à l'inverse, lorsque s est supérieur à la valeur seuil, celui-ci préfère coopérer.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Un résultat intéressant du modèle, bien que relativement intuitif, est que le seuil d'intensité du sentiment de honte, qui doit être dépassé pour assurer la coopération, dépend négativement du niveau d'exposition du comportement du deuxième joueur. En d'autres termes, un même joueur 2 peut choisir de dévier lorsque le niveau d'exposition de son comportement est minimal (le coût émotionnel à dévier étant faible dans ces circonstances), tout en préférant coopérer lorsque qu'il joue avec un niveau maximal d'exposition (le coût émotionnel devenant trop important dans cet autre contexte). Ainsi, le rôle positif de la honte, dans la promotion de la coopération, est renforcé à mesure que le comportement des joueurs peut être observé avec précision.

Avec cette modélisation, on voit à nouveau comment la prise en compte explicite et spécifique d'émotions particulières offre un niveau de nuance supérieur permettant d'apporter une réponse théorique à l'émergence grandissante de résultats expérimentaux contredisant le modèle de représentation classique. Ainsi, le modèle de Tadelis est en mesure d'expliquer théoriquement les résultats obtenus par Dana *et al.* (2006), en ce qui concerne le rôle du niveau d'exposition des choix stratégiques des agents dans la mise en œuvre d'une conduite morale.

En outre, bien que les émotions aient pu être considérées, dans le passé, comme nuisant à l'expression des capacités cognitives des agents, les empêchant de prendre des décisions de manière rationnelle, l'intégration des émotions dans les modèles de prise de décision ne s'effectue pas au sacrifice d'une remise en question de la rationalité des agents. Tadelis (2008) explique en effet que les joueurs 1 sont capables d'anticiper la honte que ressentiraient les joueurs 2 en déviant et que cela influe sur leur choix stratégique de faire confiance ou non à la première étape. Ainsi, il montre formellement mais aussi à l'aide d'une expérimentation que les joueurs 1 sont significativement plus nombreux à faire confiance, lorsque les partenaires jouent avec le niveau maximal d'exposition, comparativement au cas où le niveau d'exposition est minimal.

Les joueurs 1 utilisent donc de manière rationnelle leurs croyances sur le type de joueur 2 avec lequel ils jouent et sur la honte que celui-ci est amené à ressentir, conditionnellement au contexte d'émergence. Le travail de Tadelis montre, par conséquent, comment la rationalité au sens de l'économiste classique, peut s'adapter à l'intégration des émotions morales et en particulier de la honte.

3.3.1.3. La place de la gratitude dans le champ de l'économie comportementale

Dans la littérature, seuls quelques modèles ont examiné le rôle économique de la gratitude, et ce nombre est encore plus faible s'agissant des travaux qui ont cherché à intégrer (à la fois volontairement et de manière explicite) cette émotion spécifique comme un élément clé de la formalisation⁹.

9. De nombreux modèles de réciprocité englobent implicitement le rôle de la gratitude, mais le manque de détails, relativement aux mécanismes psychologiques sous-jacents, inhérents à l'analyse du rôle économique des émotions, implique qu'ils ne puissent être considérés comme des modèles portant sur la gratitude.

3.3.1.3.1. La gratitude représentée comme un déterminant du don

L'un d'entre eux est l'œuvre de J. Atsu Amegashie en 2006. Ce travail, qui fait partie de la littérature philanthropique, met l'accent sur les déterminants des comportements de don dans le jeu du dictateur¹⁰ et vise à ouvrir la boîte noire de ce qu'Andreoni (1989,1990) a appelé « le goût de donner » (de l'anglais « warm-glow giving », notre traduction). En effet, Amegashie (2006, p. 10, notre traduction) explique : « Andreoni avance que les individus donnent parfois dans le but de ressentir « le goût de donner ». Une composante de ce « goût de donner » est peut-être la gratitude puisque les individus pourraient donner en vue de recevoir la gratitude [ressentie et exprimée par celui à qui ils ont donné]. En un mot, nous sommes capable de rationaliser le ressenti du « goût de donner » en arguant que les donateurs dérivent de l'utilité à partir de l'acte de donner *per se* précisément parce que les bénéficiaires dérivent de l'utilité de l'acte de recevoir *per se* ».

Formellement, Amegashie (2006) définit une fonction d'utilité différente pour chaque type de joueur impliqué dans le jeu du dictateur, à savoir le receveur et le donneur (ou dictateur).

La fonction d'utilité du destinataire est composée de deux arguments. D'une part, il tire de l'utilité de la consommation matérielle du don qu'il reçoit dans le jeu (notée t). D'autre part, une certaine utilité découle également de l'émergence de la gratitude chez lui. Cette retombée émotionnelle est conçue, dans le modèle, comme une fonction positive du sacrifice (en proportion) effectué par le dictateur à travers son acte de don ($\frac{t}{m}$, où m désigne la dotation initiale du dictateur). Dans l'article, la fonction d'utilité du destinataire qui est utilisée dans l'analyse est définie comme suit :

$$U_R = \ln(t) + \alpha \ln\left(\frac{t}{m}\right)$$

Où $\alpha > 0$.

Le donateur, qui est par ailleurs le seul joueur à prendre une décision stratégique dans le jeu, est également représenté avec une fonction d'utilité constituée de deux arguments. Tout d'abord, nous retrouvons l'utilité que le donneur dérive de la consommation matérielle du reliquat de sa dotation initiale ($c = m - t$). Le second argument n'est autre que la fonction d'utilité même du destinataire. Ainsi, cette spécification implique que, outre la consommation égoïste du capital résiduel au don, le bienfaiteur apprécie également l'utilité qu'il est capable de générer et de percevoir chez l'autre, que ce soit en termes de consommation matérielle ou de sentiments de gratitude (qui se trouve être à la source du « goût de donner »).

10. Dans lequel deux individus, le bienfaiteur (ou donneur, dans l'article) et le destinataire, jouent ensemble. Au début du jeu, le donneur est doté d'une certaine quantité de ressources (généralement des unités d'argent expérimental) et doit décider d'une répartition de cette dotation entre lui-même et l'autre joueur. Pour sa part, le destinataire n'a pas de choix stratégique à faire, il est soumis à la décision du donneur.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

La fonction d'utilité du donneur qui est utilisée dans le papier est définie de la manière suivante :

$$U_D = \ln(c) + \beta U_R$$

Où $0 < \beta < 1$.

Le modèle est tout simplement résolu en maximisant l'utilité du donneur par rapport au niveau de don t . De manière assez surprenante, l'auteur a montré que, pour certaines valeurs de α et β , le donneur peut être disposé à donner plus de la moitié de sa dotation, alors même que celui-ci accorde plus de poids à sa propre consommation matérielle qu'à l'utilité du destinataire du don ($0 < \beta < 1$). Pour que cela soit possible, le destinataire doit être très reconnaissant (ce qui implique que la gratitude doit être intensément ressentie, ainsi qu'exprimée avec enthousiasme) et, dans le même temps, le donneur doit avoir un « goût de donner » suffisamment prononcé (c'est-à-dire que, α et β doivent être suffisamment grands).

Plus généralement, ce modèle introduit un élément crucial en ce qui concerne la mise en place de l'influence de la gratitude. En effet, l'émotion de gratitude, ainsi que les chaînes de réciprocité positive, doivent être suscitées et initiées pour rendre les avantages de leur influence disponibles. Ainsi, on peut se demander ce qui est à l'origine de la motivation à réaliser le premier pas. Un motif possible à agir généreusement est supposé être l'expérience anticipée de sentiments agréables liés au « goût de donner », qui est ici conçue comme la valorisation émotionnelle de l'augmentation de l'utilité que l'on peut causer à autrui, à travers l'acte de donner. En outre, selon Amegashie (2006), la détection de l'émergence de la gratitude, et son anticipation, chez le partenaire, fait partie de ce qui touche un individu au cours de la réalisation d'un don ¹¹.

La capacité de détecter et d'anticiper l'émergence d'émotions chez les autres fait écho avec le processus émotionnel de l'empathie. Il est intéressant de noter qu'Amegashie (2006) considère que l'on a probablement une capacité d'empathie d'autant plus grande s'agissant d'individus géographiquement et socialement proches de nous. Cette évaluation présente le lien entre l'influence de la reconnaissance et la proximité entre les agents dans les relations sous un autre angle.

A la lumière de cette réflexion, nous sommes amenés à développer notre compréhension de l'influence dynamique de la gratitude sous un nouveau jour. En effet, ce processus pourrait prendre la forme suivante : la proximité entre deux individus augmente leur capacité à faire preuve d'empathie l'un envers l'autre, ce qui facilite à la fois l'anticipation et la détection de l'émergence de la gratitude chez l'autre, qui participe à développer le « goût

11. Une autre explication possible concernant la motivation à agir de manière prosociale en premier, peut être l'anticipation stratégique des actes de réciprocité positive potentiels que cela peut inciter chez le partenaire, à l'avenir. Néanmoins, comme mentionné plus haut, il semble que ce type de comportement stratégique, lorsqu'il est saillant, ne parvienne pas à déclencher l'émergence de la gratitude (mais plutôt un sentiment d'endettement), ce qui ne permet pas d'inciter à l'adoption d'un comportement réciproque, avec la même intensité. Ceci tend à dissuader l'utilisation de telles stratégies, lorsque l'on a pas d'intention sincère d'entreprendre des actions bienveillantes, à moins que celles-ci ne soient mises en œuvre de manière suffisamment trompeuse.

de donner ». Cette anticipation peut se transformer en une incitation à la réalisation de comportements bienveillants de la part d'un agent et l'exécution effective de tels actes peut finalement conduire à l'émergence de la gratitude chez l'autre, probablement suivie par son expression (allant d'un simple « merci » à des actes de réciprocité positive tels que des contre-dons). En fin de compte, la réalisation de ces interactions positives tend à augmenter le niveau de proximité entre les deux agents, ainsi que la probabilité qu'ils puissent bénéficier d'autres cycles d'interactions de ce type, à l'avenir.

Pour autant, on peut se demander ce qui détermine initialement le niveau de proximité entre deux individus, car cela semble être un facteur décisif dans le fait d'être à l'initiative, en entreprenant des actions prosociales pouvant aboutir à des spirales de réciprocité positive, voire à des effets d'entraînement au sein des groupes.

Nous pensons que les agents possèdent une disposition naturelle à se sentir plus ou moins proches d'individus anonymes, lorsqu'ils les rencontrent pour la première fois. Bien que notre modèle ne cherche pas à répondre à cette question particulière, nous fournissons ici quelques éléments de réflexions sur cette question.

Selon nous, la façon la plus simple de conceptualiser la propension des agents à se sentir plus ou moins proche d'une personne inconnue, devrait être de la considérer comme une tendance générale (presque comme un trait de personnalité), qui s'exprimerait de la même façon, peu importe l'individu particulier auquel on est confronté.

Une autre possibilité pourrait être de penser cette disposition comme étant dépendante du contexte et, plus précisément, comme étant une fonction des informations inférées à propos de la personnalité des autres. Ainsi, il pourrait en résulter que l'on se sente plus proche de ceux chez qui l'on perçoit des caractéristiques similaires. En ce sens, le fait que les individus apprennent à se connaître pourrait être un moyen d'augmenter leur niveau de proximité les uns envers les autres, se traduisant par une plus grande probabilité d'assister au démarrage de chaînes de comportements positifs. De telles conditions constituant sans doute un terrain propice à l'émergence de la coopération, tout comme à l'installation d'un climat émotionnel positif.

A titre d'illustration, on peut penser à ce qui se passe habituellement quand une équipe est créée, ou lorsque de nouveaux membres sont ajoutés à celle-ci. La plupart du temps, un tel événement débute par une présentation des (nouveaux) membres de l'équipe. Compte tenu de ce que nous venons de voir, la mise en place de ce type de comportement semble fondé, car une augmentation du niveau de proximité entre les agents peut se produire au sein de toute l'équipe, à mesure que les membres sont introduits, probablement en raison d'une amélioration de la capacité des agents à être empathiques les uns envers les autres. Ces derniers éléments sont cohérents avec plusieurs résultats empiriques comme, par exemple, le fait que l'identification des partenaires, avant de participer à des jeux de dilemmes sociaux, permette de réduire la distance sociale et d'augmenter le niveau de coopération (Andreoni et Petrie, 2004).

Enfin, si l'on change de perspective, on peut également tenir compte du fait que les êtres humains sont, à l'évidence, capables de manipuler les indices servant, au cours d'une première rencontre, à l'extrapolation d'informations sur leur personnalité. Que ce soit à

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

travers les vêtements que l'on porte (aux côtés des autres dimensions gérables de l'apparence physique), la façon dont on se déplace, dont on parle, dont on réagit, ou par l'intermédiaire de ce que l'on dit, les émotions que l'on affiche (même si les émotions peuvent être difficiles à simuler) etc. ; nous avons une capacité de contrôle sur l'image personnelle que l'on renvoie, éventuellement dans le but d'attirer la sympathie des autres. Cette image personnelle peut aller d'une authentique représentation de la véritable personnalité d'un individu à un portrait trompeur servant à des fins manipulatoires. Toutefois, cette capacité de contrôle sur l'image pourrait être un facteur important dans la capacité d'un individu à faire en sorte que les autres se sentent proches de lui, ce qui l'aiderait à obtenir de l'aide ou d'autres avantages auprès de ceux qui l'entourent.

En fin de compte, on pourrait très bien imaginer que le niveau réel de proximité qu'un agent éprouve à l'égard d'un individu anonyme puisse se déterminer en réponse à la question suivante : dans quelle mesure, sur la base des indices disponibles, les traits de personnalité, qui caractérisent apparemment l'autre, sont appréciés par l'agent ?

En outre, il est également possible que l'ensemble des traits de personnalité que l'on apprécie chez les autres soient déterminés par des processus favorisant l'appréciation des traits de personnalité qui nous caractérisent.

Bien que très intéressant et novateur, le modèle d'Amegashie (2006) possède également certaines limites. En ce qui concerne la représentation du rôle économique de la gratitude, sa limite prépondérante est, à notre avis, le fait que la formalisation soit centrée sur le jeu du dictateur. Ce jeu est en effet très limité dans la mesure où il n'y a qu'un seul joueur qui peut prendre une décision économique, l'autre agent ne faisant que subir le choix du premier. Il n'y a donc pas, à proprement parler, d'interaction stratégique entre les individus dans cette situation et la pertinence d'un tel jeu a même parfois été interrogée par certains économistes (Guala et Mittone, 2010).

Ce choix est clairement compréhensible étant donné l'objectif de l'auteur dans cet article, qui est d'analyser les déterminants du don. Néanmoins, comme nous l'avons vu, le rôle économique de la gratitude va au-delà des questions de réciprocité positive et de l'anticipation du plaisir dû au « goût de donner », qui accompagne la perception de sentiments de gratitude chez le bénéficiaire d'un don. En ce sens, le jeu du dictateur ne constitue pas un cadre approprié pour analyser les nombreuses caractéristiques de cette émotion cruciale. En effet, il ne donne aucune possibilité d'un quelconque acte de réciprocité de la part du bénéficiaire (mis à part le niveau de reconnaissance qu'il ressent et exprime) et, par ailleurs, étant donné que le jeu est analysé dans sa version en un coup, l'intégration de l'aspect dynamique de l'influence de la gratitude est, de fait, exclue. Enfin, le jeu du dictateur, tel qu'il est présenté dans ce modèle (qui est conforme à la présentation standard du jeu), fait intervenir deux joueurs seulement, ce qui est regrettable à l'égard de l'influence de la gratitude au niveau du groupe, ainsi que, vis-à-vis de la capacité de la gratitude à se propager dans un réseau d'agents.

3.3.1.3.2. La diffusion de la coopération dans un réseau d'agents par la réciprocité positive directe et indirecte

Vis-à-vis de la construction d'un réseau d'agents, il existe également, dans la littérature évolutionniste en économie, une analyse formelle du rôle de la gratitude. En particulier, Nowak et Roch (2007) ont présenté un modèle, basé sur les marches aléatoires, qui remet en question le rôle adaptatif de la réciprocité indirecte. Dans cet article, les auteurs considèrent un ensemble d'agents économiques, définis sur la base de stratégies concernant l'initiation et la perpétuation, par renvoi direct ou par transmission à un tiers, d'une chaîne d'actes altruistes, au sein du réseau qu'ils constituent.

Dans la situation la plus simple, chaque agent est caractérisé par une probabilité q d'initier la chaîne, et par une probabilité p de perpétuer celle-ci, sachant que, chaque fois que la chaîne passe par un joueur, celui-ci bénéficie d'un acte prosocial de valeur b (de la part de l'agent précédent, qui perpétue la chaîne) et que chaque fois qu'un agent passe la chaîne, il supporte un coût c (représentant les ressources utilisées pour fournir un acte altruiste à l'agent suivant), avec $b > c$. Le perpétuement de la chaîne représente ainsi le fonctionnement de la réciprocité positive indirecte, puisqu'il correspond à une situation où une personne reçoit d'abord de l'aide, avant d'aider quelqu'un d'autre à son tour.

Plus tard, les auteurs considèrent également une situation plus complexe dans laquelle les agents ont également la possibilité de faire preuve de réciprocité positive directe. Ainsi, en plus de leur probabilité d'initier et de transmettre la chaîne à un tiers, les agents sont également définis à l'aide de r , qui désigne la probabilité avec laquelle ils renvoient directement l'acte altruiste qu'ils viennent de recevoir à leur bienfaiteur.

Pour étudier la valeur sélective de la réciprocité positive indirecte, Nowak et Roch (2007) analysent la mise en concurrence de diverses stratégies, opposées deux à deux, et différenciées par leur degré « d'aptitude à coopérer », définie à partir des probabilités q , p , et, le cas échéant, r (étant donné que des valeurs plus grandes de ces probabilités se traduisent par une plus grande « aptitude à coopérer » de la stratégie), au sein d'une population d'agents, composée de deux sous-ensembles, adoptant chacun une stratégie particulière. La plupart du temps, une stratégie plus ou moins coopérative est évaluée contre la stratégie de défection, correspondant à $q = 0$, $p = 0$ et $r = 0$, lorsque la réciprocité directe est autorisée. Un processus dynamique garantit que les proportions des deux sous-ensembles, dans la population, évoluent en faveur de celui dont la stratégie est la plus performante, en termes d'utilité.

Le principal résultat mis en évidence dans cet article est que la réciprocité positive indirecte seule ne facilite pas l'évolution de la coopération, cependant, elle peut évoluer en tant que produit dérivé de la réciprocité directe. En effet, lorsque la réciprocité directe n'est pas autorisée, les résultats montrent que la stratégie avec « l'aptitude à coopérer » la moins importante supplante toujours l'autre. Ainsi, la stratégie de défection domine logiquement toutes les stratégies qui ont une « aptitude à coopérer » positive. A l'inverse, lorsque les agents ont la possibilité d'adopter un comportement de réciprocité directe, les résultats indiquent qu'une stratégie, $S(p, q, r)$, est évolutivement stable face à l'invasion par une stratégie dont « l'aptitude à coopérer » est inférieure, lorsque la valeur de r dépasse $\frac{c}{b}$, le rapport coût-bénéfice de l'acte altruiste.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Ce résultat paraît logique dans la mesure où, plus la probabilité est grande que la chaîne d'actes altruistes nous revienne, après que l'on l'ait perpétuée, plus on est intéressé par le fait que cette chaîne se poursuive. Ainsi, il est systématiquement d'autant plus pertinent de perpétuer cette chaîne (par l'adoption d'une stratégie plus coopérative) quand la probabilité qu'elle revienne directement à nous est supérieure au rapport coût-bénéfice, lequel affecte négativement l'incitation économique à coopérer.

En outre, Nowak et Roch (2007) ont également remarqué que si le paramètre r tend vers 1, alors la réciprocité positive indirecte devient moins importante pour le succès des stratégies de coopération. Néanmoins, dans la réalité, une réciprocité directe n'est pas toujours possible ou nécessaire, s'il n'y a pas d'opportunité ou de besoin d'aider le donneur immédiatement. Pour prendre en compte ce phénomène, les auteurs examinent aussi ce qui se passe quand une borne supérieure est imposée à la valeur de r . Dans ce cas, ils constatent que le processus de sélection fait continuellement pression en faveur de plus en plus de réciprocité positive indirecte, ce qui témoigne de son intervention complémentaire dans la promotion de la coopération.

Un autre résultat intéressant de ce modèle concerne la comparaison de la stratégie d'altruisme spontané, caractérisée par des valeurs élevées du paramètre q , avec la stratégie d'altruisme liée à la gratitude, caractérisée par des valeurs élevées du paramètre p . Dans les faits, il a été démontré que l'altruisme lié à la gratitude peut conduire à des niveaux beaucoup plus élevés « d'aptitude à coopérer ». En outre, si $r > \frac{c}{b}$, les « coopérateurs spontanés » (pour lesquels $p = 0$) sont surpassés par les agents dont « l'aptitude à coopérer » est liée à la gratitude ($p > 0$).

Selon Nowak et Roch (2007, p. 608, notre traduction) : « La gratitude et d'autres émotions positives, qui encouragent la propension à aider (étant donné le paramètre p dans notre modèle), peuvent évoluer par la sélection naturelle. Il est également concevable que la réciprocité [indirecte] en amont soit un sous-produit d'émotions qui ont avant tout évoluées afin de faciliter la réciprocité directe ».

Pour nous, la présence d'actes de réciprocité positive indirecte, dans la vie quotidienne, est une marque de la valeur sélective du sentiment de gratitude. En effet, il semble que l'évolution ait favorisé la survie d'individus pour qui les sentiments reconnaissants sont si puissants que, lorsque l'acte à l'origine de leur émergence ne peut être rendu au bienfaiteur initial (par exemple parce qu'il n'est pas immédiatement dans le besoin), l'envie d'agir de manière prosociale finit par se traduire par la réalisation d'un comportement bienveillant à l'encontre d'un tiers, en formant ainsi une chaîne d'actes altruistes. S'il en avait été autrement, ces comportements auraient disparu au profit d'autres, plus égoïstes. Or, au contraire, il semble que les groupes dans lesquels les chaînes de comportements bienveillants peuvent durer plus longtemps, sont capables de supplanter ceux qui sont incapables de maintenir de telles chaînes de comportements coopératifs.

3.3. *L'intégration explicite des émotions morales dans les modèles de prise de décision économique*

Ce type de modélisation permet de réfléchir sur le rôle adaptatif de la gratitude et offre un possible moyen d'expliquer comment cette émotion peut se propager au sein des groupes. Dans le travail que nous venons de voir, les auteurs examinent à la fois comment et quand le fait d'être reconnaissant, dans le sens de perpétuer (directement ou indirectement) une chaîne d'actes altruistes, renforce l'aptitude à la survie, et la façon dont le processus de sélection dynamique des stratégies favorise ou restreint l'utilisation de tels comportements coopératifs.

Néanmoins, dans les modèles évolutionnistes, le seul choix stratégique des agents se résume à la fixation de probabilités d'occurrence de différents comportements possibles et, en conséquence, les comportements effectifs sont considérés comme étant exécutés au hasard. Dans ce travail, nous sommes également intéressés par l'analyse du rôle de la gratitude au sein d'un groupe (ou d'une population) d'agents, ainsi que par la façon dont cette influence peut évoluer dynamiquement. Cependant, nous préférons envisager les décisions effectives des agents comme résultant de la maximisation de leur utilité, plutôt que comme étant l'expression d'une probabilité fixée stratégiquement au préalable. Ainsi, le cadre évolutionniste ne nous apparaît pas être le plus approprié à la prise en compte de l'influence des émotions sur le processus de décision.

L'ensemble des travaux que nous venons de revoir se concentre sur le rôle spécifique d'une émotion particulière. Ceci nous semble être un travail nécessaire dans l'appréhension du fonctionnement de l'influence d'une émotion. Toutefois, nous considérons que les émotions morales forment un ensemble d'outils et que celles-ci ont vocation à être complémentaires et à pouvoir toutes participer à la prise de décision, lors d'une même situation, en fonction du contexte spécifique et du déroulement effectif d'une interaction. En ce sens, d'autres modélisations, ayant vu le jour ces dernières années, sont porteuses d'idées originales dans l'optique d'une modélisation plus fidèle du rôle des normes et de la variété d'émotions morales influant conjointement sur le processus de décision.

3.3.2. Vers des modélisations originales intégrant plusieurs émotions morales

3.3.2.1. Le rôle des émotions prosociales dans le modèle de Bowles et Gintis (2005)

Parmi ces travaux, on retrouve le modèle de Bowles et Gintis (2005), qui s'inscrit dans le cadre d'un jeu du bien public avec une phase de punition, soit une phase subséquente aux décisions de contributions, durant laquelle les participants ont le droit d'infliger une sanction à ceux qui jouent avec eux. Une telle présentation du jeu permet de retranscrire les représailles qui sont susceptibles d'être mises en œuvre, dans la vie quotidienne, lorsque, par exemple, un individu adopte un comportement de passager clandestin lors d'un travail de groupe (qu'ils s'agissent de réprimandes orales, de violences physiques, de mises à l'écart etc.).

Formellement, en considérant d'abord le cas de $n = 2$ agents participant à un jeu du bien public, avec phase de punition, joué en un coup, on désigne l'ensemble des joueurs par $I = \{i, j\}$ avec $i \neq j$.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Au début du jeu, chaque joueur se voit attribuer une dotation personnelle égale à 1. Par la suite, chacun d'eux doit décider s'il souhaite garder ce capital pour lui-même ou bien investir tout ou partie de ses ressources dans un bien public qui bénéficie à tous les membres du groupe. Dans la pratique, les joueurs déterminent simultanément le montant de contribution $a_i \in [0, 1]$ ($i \in I$) qu'ils désirent investir dans un bien public rétribuant la somme $q \sum_i a_i$ à chacun d'entre eux, avec $0.5 < q < 1$, ce qui assure au jeu d'avoir les caractéristiques d'un dilemme social. En effet, dans ces conditions, la stratégie dominante d'un joueur égoïste est de ne rien contribuer, même si, collectivement, l'optimum est atteint lorsque les deux joueurs contribuent $a_i = 1$.

Durant la phase suivante, après avoir été informé du niveau de contribution de l'autre joueur, chaque agent a la possibilité d'infliger, à ses frais, une punition matérielle à son partenaire. Formellement, chaque agent $i \in I$, peut imposer une sanction μ_{ij} ($i \neq j$) à l'autre joueur, au coût $c(\mu_{ij})$. A titre d'exemple illustratif, les auteurs définissent, dans l'article, la fonction de coût suivante : $c(\mu_{ij}) = \frac{\gamma}{2}\mu_{ij}^2$. Cette fonction se rapproche notamment du protocole expérimental utilisé par Fehr et Gächter (2000), pour ce qui est de la mise en œuvre des punitions. Ainsi, l'expression retenue par les auteurs pour représenter le gain final du joueur $i \in I$, $i \neq j$, dans le jeu, est :

$$\pi_i = 1 - y_i + a(y_i + y_j) - \mu_{ji} \quad (3.3.4)$$

Il convient de noter que, pour des raisons expliquées ci-dessous, les auteurs n'intègrent pas le coût de la punition dans l'expression du gain matériel des agents, bien que ce coût soit, bien entendu, présent dans la fonction d'utilité que nous présentons en détails ci-après.

L'intérêt de Bowles et Gintis (2005) pour un tel jeu a été suscité par un résultat particulièrement notable (obtenu par le traitement de données expérimentales) : répondre à la punition, en augmentant son niveau de contribution au bien public, ne peut pas être considéré comme une meilleure réponse, s'agissant de la maximisation des gains d'un joueur dans le jeu. Autrement dit, même en tenant compte du coût lié au fait d'être puni par les autres, il est toujours matériellement rentable d'avoir un comportement de défection. En outre, il a également été montré que les sanctions non monétaires (comme les critiques, par exemple) sont également efficaces pour prévenir des comportements de « free-riding » (Maslet, Noussair, Tucker et Villeval 2003). Dans ces circonstances, étant donné que les résultats empiriques tendent à prouver que la punition est en mesure de favoriser les comportements de contribution (Fehr et Gächter, 2000), leur influence ne peut pas être expliquée par la seule prise en compte des gains matériels. Ainsi, d'autres explications doivent être proposées que la crainte de subir un coût matériel, en étant puni, pour interpréter l'efficacité des sanctions dans la dissuasion des comportements égoïstes, dans le contexte de dilemmes sociaux tel que le jeu du bien public.

Par conséquent, les auteurs suggèrent que, lorsque des punitions matérielles sont employées, au-delà de leur impact tangible, l'effet complémentaire du sentiment de honte entre également en jeu. En recourant aux travaux d'Elster (1998), ils expliquent que les

3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique

sanctions matérielles (ainsi que probablement les non matérielles) véhiculent des émotions négatives comme le mépris (auquel on peut également ajouter la juste colère), qui sont des déclencheurs directs de la honte.

Ainsi, en résumé, on suppose ici que le fait de répondre à la crainte d'être puni par l'adoption d'un comportement (suffisamment) coopératif, est d'une part un moyen d'éviter le coût matériel de la punition, mais, d'autre part, il s'agit surtout d'un moyen d'échapper aux conséquences émotionnelles d'être (publiquement) réprimandé, soit ressentir le sentiment de honte.

Les auteurs abordent également le rôle de l'émotion de culpabilité, qui est supposée émerger lorsqu'un individu agit en contradiction avec la norme de comportement qu'il a intériorisée comme étant moralement et socialement adaptée, relativement à une situation particulière.

En somme, l'émergence de la culpabilité, ainsi que de la honte, qui survient après qu'un autre individu ait fait part de ses sentiments négatifs (colère et le mépris), par la punition, représentent des forces essentielles, supposées orienter la conduite des agents vers le respect des normes qu'ils ont intériorisées. En références aux travaux d'Antonio Damasio (1994), les émotions (négatives) sont ici conçues comme des marqueurs somatiques qui influencent le processus de décision d'une manière comparable à la douleur, qui est décrite par Bowles et Gintis (2005) comme une sensation très désagréable ne pouvant être ignorée et que l'on cherche à tout prix à éviter. En d'autres mots, l'anticipation d'états émotionnels désagréables, déjà expérimentés par le passé, informe les agents des dangers potentiels de leurs comportements égoïstes.

En pratique, le modèle de Bowles et Gintis (2005), tout en s'appuyant sur divers précédents travaux de formalisation (Geanakoplos, Pearce et Stacchetti, 1989 ; Falk et Fischbacher, 2006 (le travail original était déjà disponible en 1998, bien que non publié) ; Levine, 1998 ; et Sethi et Somanathan, 2001), propose une fonction d'utilité originale, qui saisit cinq objectifs différents, que chaque individu cherche à atteindre pour maximiser son niveau de satisfaction : (1) ses gains matériels individuels ; la façon dont on valorise les gains obtenus par les autres joueurs, qui dépend à la fois de (2) son niveau d'altruisme et de (3) sa sensibilité à l'égard des questions de réciprocité ; et, enfin, l'intensité avec laquelle il se sent (4) coupable et (5) honteux, en réponse à ses propres actions, ainsi qu'aux actions de rétorsion que son comportement induit chez les autres (sanctions).

Ainsi, en plus de leurs propres gains dans le jeu, les individus se préoccupent également du sort des autres, ce qui se traduit formellement par le fait que l'agent $i \in I$ accorde un certain poids aux gains matériels de l'autre joueur π_j . Ce poids est défini comme suit :

$$\delta_{ij} = \alpha_i + \lambda_i(a_j - a_i^*) \quad (3.3.5)$$

Où a_i^* désigne la valeur de la norme de contribution qui a été intériorisée par le joueur i et α_i représente la préoccupation altruiste inconditionnelle dont le joueur $i \in I$ fait preuve à l'égard de l'autre agent. Si $\alpha_i \geq 0$, i peut être qualifié de bienveillant car cela

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

implique qu'il démontre une tendance inconditionnelle à se soucier des gains matériels de l'autre agent, dans le jeu. Au contraire, si $\delta_i < 0$, i peut être vu comme une personne malveillante, affichant une tendance inconditionnelle à valoriser négativement le sort de j , dans le jeu. Dans l'article, seul le cas $\delta_i \geq 0$ est examiné, bien que les deux possibilités puissent être envisageables.

Au-delà de cette préoccupation altruiste inconditionnelle, le poids accordé aux gains matériels de l'autre dépend aussi de questions de réciprocité. En effet, les agents sont supposés être des « strong reciprocators » (coopérateurs conditionnels), dans le sens où ils sont caractérisés par une prédisposition (plus ou moins ancrée en eux-mêmes, selon la valeur de λ) à être prêt à coopérer avec ceux qui en font de même et à vouloir se venger des individus non-coopératifs (par le biais de défections, si ce n'est par la punition), même si ce comportement ne peut être justifié en termes de leur intérêt personnel, sur le plan matériel.

Formellement, le paramètre $\lambda_i \geq 0$, qui peut être considéré comme la sensibilité de l'agent i aux questions de réciprocité, vient multiplier la différence entre la contribution de j et la norme de i . Il convient de noter que cette différence peut être considérée comme un indice utilisé par i afin d'évaluer le niveau d'aptitude à coopérer de j , sur la base du choix stratégique mis en œuvre par ce dernier.

Par conséquent, dans la pratique, lorsque $a_j > a_i^*$ (et si $\lambda_i \neq 0$), j est évalué comme coopératif par l'agent i . En outre, plus la contribution de j est grande, en comparaison avec la norme de i , plus celui-ci valorise le sort de j , ce qui le pousse à être coopératif en retour.

A l'inverse, lorsque $a_j < a_i^*$ (et si $\lambda_i \neq 0$), l'agent i perçoit le joueur j comme un individu non-coopératif. De plus, moins la contribution de j est importante relativement à la norme de i , moins celui-ci valorise les gains de j , ce qui le motive à se venger.

Le coût à supporter pour punir l'autre joueur n'est pas intégré dans l'expression dans π_i ($i \in I$) afin d'éviter que le poids qui est placé sur les gains matériels de l'autre joueur ne soit appliqué à ce coût. Dans le cas contraire, si on suppose que i se soucie du sort de j ($\delta_{ij} > 0$), celui-ci pourrait être disposé à adopter un comportement plus coopératif, de peur que j soit obligé de payer pour le punir, s'il ne contribue pas assez, ce qui est considéré comme peu plausible par les auteurs.

De plus, les individus sont également conçus comme averses à la violation de leur propre norme de contribution. En conséquence, il est supposé dans le modèle, que chaque joueur $i \in I$ souffre d'un coût psychologique de niveau $\beta_i(a_i^* - a_i)^2$, du fait de l'émergence de la culpabilité, lorsque sa contribution y_i diffère de la norme \bar{y}_i que celui-ci a intériorisée. Dans le papier, pour tout $i \in I$, la valeur de la norme de contribution a_i^* est conçue comme une fonction de la dotation du joueur. Néanmoins, ce point particulier n'est y pas développé en détail.

En raison de la présence d'un carré, les agents supportent un coût lié à la culpabilité aussi bien lorsque leur niveau de contribution ne parvient pas à atteindre celui de leur norme, que lorsqu'ils dépassent la valeur de celle-ci.

Lorsque $a_i < a_i^*$, le coût psychologique est facilement interprétable. En effet, la culpabilité est un sentiment qui intervient lorsque l'on viole une norme sociale ou morale que l'on considère comme sensée, en particulier dans le cas où cette violation représente un heurt à l'encontre d'autres personnes. Ici, quand i ne parvient pas à atteindre le niveau de sa contribution normative, la norme est effectivement violée et, dans le même temps, cela est nuisible pour l'autre agent puisque un moindre montant de contribution, de la part de l'agent i , implique un moindre bénéfice, issu du bien public, pour chacun des agents.

Cependant, quand $a_i > a_i^*$, l'interprétation est moins intuitive. Par exemple, on pourrait penser que les individus puissent se sentir fiers d'agir d'une façon meilleure encore que ce qui est recommandé par leur norme. Sur ce point, l'explication fournie par Bowles et Gintis (2005) consiste à considérer que les agents font face à plusieurs normes qu'ils doivent respecter, dans divers dilemmes. Ainsi, dépasser le niveau de leur norme, dans une situation particulière, vis-à-vis d'un dilemme spécifique, peut les empêcher de respecter leur norme de comportement dans au moins une des autres situations de dilemme dans lesquelles ils sont impliqués, d'où l'émergence de la culpabilité.

Enfin, la valeur du paramètre β_i exprime le niveau individuel de l'intensité avec laquelle l'agent $i \in I$ se sent coupable, pour des valeurs données de a_i et a_i^* .

Par ailleurs, en plus du coût matériel lié au fait d'être puni, les auteurs incluent également, dans la fonction d'utilité, le coût *psychologique* lié au fait d'être puni s_i ($i \in I$), résultant de la sensation du sentiment de honte. s_i est défini pour tout $i \in I$ et $i \neq j$ comme étant égal à :

$$s_i = \sigma_i(a_i^* - a_i)\mu_{ji} \quad (3.3.6)$$

Avec σ_i , la sensibilité de l'agent i à l'émergence de la honte.

Il est intéressant de noter que l'impact de la honte sur l'agent qui est puni varie en fonction de son propre comportement relativement à la norme qu'il a intériorisée. En effet, s_i est positif seulement quand $a_i < a_i^*$. Dans ce cas, l'augmentation de son niveau de contribution permet à l'agent i de diminuer le coût psychologique lié au sentiment de honte. Au contraire, lorsque $a_i > a_i^*$, s_i représente l'influence de la rancune puisque l'agent i se sent contrarié d'être puni alors même que, selon lui, son comportement n'est pas considéré comme répréhensible, ce qui se traduit par une moindre volonté de contribuer.

Par conséquent, l'émergence des émotions morales dépend, dans ce modèle, des normes qui sont intériorisées par les agents, ce qui confère à ces dernières une place de choix dans la modélisation. Tôt dans le papier, Bowles et Gintis (2005) font référence à Kenneth Arrow (1971). Celui-ci conçoit les normes morales et sociales comme des réactions possibles, de la société, à des défaillances du marché, qui résultent notamment d'un manque de confiance entre les agents, dans des situations nécessitant une entente coopérative (comme les dilemmes sociaux). Pour justifier pourquoi ils préfèrent considérer les normes comme

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

étant internalisées par les agents (à savoir intégrées dans leur fonction objectif) plutôt que comme de simples contraintes, les auteurs expliquent que les sociétés humaines sont très complexes et qu'il est très difficile de mesurer, pour toutes les normes qui composent ces sociétés, les bénéfices et coûts qui sont liés au fait de se conformer, ou au contraire de violer, toutes ces normes. Ainsi, réaliser une optimisation à grande échelle (soit en considérant une multitude de situations à la fois) en considérant les normes comme des contraintes peut s'avérer être excessivement source d'erreurs pour les agents. Face à cela, les auteurs soutiennent que l'internalisation des normes permet aux agents d'éliminer bon nombre de calculs de coût/bénéfice, en les remplaçant par des lignes directrices morales pour le choix de leurs actions. Internaliser les normes permet par conséquent aux agents d'être mieux adaptés que les autres, sur le plan biologique, de sorte à ce que les mécanismes psychologiques d'internalisation aient été sélectionnés par l'évolution.

Il convient aussi de souligner le fait que ce modèle prend en compte un aspect crucial du processus de décision humain, que nous avons déjà mentionné dans ce chapitre, à savoir le désir de se conformer aux attentes des autres. En effet, la norme que les agents ont intériorisée, qui intervient de diverses façons durant la prise de décision (en particulier en ce qui concerne l'influence de la culpabilité et de la honte), représente leur propre croyance à propos de ce que les autres doivent attendre d'eux.

Néanmoins, la façon dont les attentes sont intégrées dans l'émergence des émotions auto-conscientes, rend compte d'agents assez étroits d'esprit, puisqu'ils ne tiennent pas vraiment compte de l'opinion des autres sur ce qu'est le comportement normatif (ce qui correspond davantage aux véritables attentes particulières des autres). L'influence de ces émotions les pousse plutôt à respecter ce qu'ils considèrent être le comportement qui est attendu d'eux socialement, mais d'un point de vue plus global, à l'échelle de la société, sans tenir compte des individus particuliers auxquels ils font face dans le jeu.

Cette conception renvoie à l'idée d'agents déterminés, ayant de fortes convictions et qui ne se laissent pas influencés par l'opinion des autres au moment d'analyser la pertinence de leur comportement. Il semble, dans ce modèle, que de la volonté de se conformer aux attentes des autres individus soit en quelque sorte limitée par leur propre croyance sur ce que ces attentes sont ou devraient être. Toutefois, le point de vue des autres individus est finalement intégré dans la décision des agents lors de l'interaction qui a lieu entre eux, notamment par le biais de l'attention qu'ils accordent aux considérations de réciprocité, vis-à-vis desquelles les différentes conceptions de la norme (au sein du groupe) entrent en confrontation.

Tous les composants de la fonction d'utilité ayant été exposés, nous pouvons maintenant la présenter entièrement. Pour tout $i \in I$ et $i \neq j$, on a :

$$U_i = \pi_i + \delta_{ij}\pi_j - \beta_i(a_i^* - a_i)^2 - \gamma \frac{\mu_{ij}^2}{2} - \sigma_i(a_i^* - a_i)\mu_{ji} \quad (3.3.7)$$

Avant de passer à la résolution du jeu, il convient de noter ici que la forme fonctionnelle utilisée dans ce modèle favorise l'expression de l'utilité des agents en termes de préférences sur les *actions*, plutôt qu'en termes de préférences sur les *issues*. En effet, l'évaluation par

3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique

i du comportement de j se fait à la lumière de la comparaison entre l'*action* de j et l'opinion de i à propos de l'*action* normative à effectuer. En outre, l'émergence et l'influence des deux émotions de culpabilité et de honte résulte également de l'analyse, par i , de la moralité de son propre comportement, en termes d'*action*.

Faire cette distinction peut sembler un peu désuet dans le cadre du jeu de bien public et, plus généralement, dans la plupart des jeux de dilemme social, car il existe un lien spécifié entre les actions et leurs conséquences, qui est (dans la plupart des cas) parfaitement connu par tous les joueurs.

Néanmoins, nous souhaitons faire valoir ici que, par souci de réalisme, les fonctions d'utilité des modèles de préférences sociales devraient s'inscrire prioritairement dans le cadre de préférences sur les actions, puisque la moralité d'une conduite n'est pas tant en rapport avec le fait que son issue puisse être considérée comme morale (vis-à-vis des parties prenantes à l'interaction), mais plutôt avec le fait que l'agent qui la met en œuvre fasse preuve, par son choix, d'une intention d'agir en cohérence avec les principes moraux en vigueur (à la fois les siens et ceux des autres). Ce point sera discuté plus longuement à la fin de ce chapitre.

En pratique, le jeu se déroule en commençant par la phase de contribution, au cours de laquelle chaque joueur $i \in I$ choisit son niveau de contribution a_i . Puis, le jeu se poursuit avec l'étape de punition, durant laquelle chaque agent détermine le niveau de sanction μ_{ij} qu'il souhaite imposer à son partenaire, l'agent j , en réponse à la contribution a_j de celui-ci.

Pour résoudre le modèle, les auteurs utilisent un raisonnement par induction à rebours, dont la première étape consiste à déterminer les *fonctions de punition* des agents, définissant, pour chacun d'eux, le niveau de punition optimal à effectuer en fonction du niveau de contribution du partenaire.

La condition de premier ordre donne, pour tout $i \in I$ et $i \neq j$:

$$\frac{\partial U_i}{\partial \mu_{ij}} = -\gamma \mu_{ij} - \delta_{ij} = 0$$

En conséquence, l'agent doit fixer le montant de sa punition au niveau qui égalise le coût marginal de la sanction (le premier terme) au bénéfice marginal de celle-ci, à savoir la valeur qu'il accorde au fait de réduire le gain de l'autre joueur (le deuxième terme).

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Il vient alors :

$$\mu_{ij} = \begin{cases} -\frac{\delta_{ij}}{\gamma} > 0 & \text{si } \delta_{ij} < 0 \\ 0 & \text{sinon} \end{cases} \quad (3.3.8)$$

Avec :

$$-\frac{\delta_{ij}}{\gamma} = \frac{-\alpha_i - \lambda_i(a_j - a_i^*)}{\gamma} = \frac{\lambda_i(a_i^* - a_j) - \alpha_i}{\gamma}$$

Il convient de remarquer que $\delta_{ij} < 0$ n'est possible que lorsque $a_j < a_i^*$. Par l'observation de la dernière expression, il est alors évident que la fonction de punition est croissante avec le degré de réciprocité λ et décroissante avec le niveau d'altruisme α .

Ensuite, chaque agent est censé être capable d'anticiper précisément la corrélation entre son propre niveau de contribution et le montant de la sanction qui en résulte, éventuellement, de la part de l'autre joueur. Ainsi, une fois que les fonctions de punition sont intégrés dans la fonction d'utilité, la deuxième étape du raisonnement par induction à rebours consiste en la maximisation de l'utilité du niveau de contribution.

La condition de premier ordre permet de définir la fonction de meilleure réponse suivante :

$$a_i = \frac{\gamma(q(1 + \delta_{ij}) + 2\beta_i a_i^* - 1) + \sigma_i(\lambda_j(a_j^* + a_i^*) - \alpha_j) + \lambda_j}{2(\beta_i \gamma + \lambda_j \sigma_i)} \quad (3.3.9)$$

Pour chaque individu, la fonction de meilleure réponse augmente avec le niveau de contribution de l'autre joueur. Par ailleurs, on peut montrer que, pour $a_i^* > a_i$, $\frac{\partial a_i}{\partial \beta_i}$ et $\frac{\partial a_i}{\partial \sigma_i}$ sont toutes deux positives, ce qui implique que plus les agents sont sensibles aux émotions de culpabilité et de honte, plus le niveau de leur contribution est important (par le déplacement vers le haut de la fonction de meilleure réponse). En d'autres termes, un changement dans la force avec laquelle les émotions de culpabilité ou de honte sont ressenties, par exemple, en considérant que l'on remplace un des agents (ou les deux), par un individu dont la personnalité est telle qu'il est sujet à ressentir de manière plus intense ces émotions, toutes choses égales par ailleurs, implique des comportements et un équilibre plus coopératifs¹².

12. Résultats que l'on retrouve empiriquement d'après Tangney, Stuewig et Mashek (2007).

3.3. L'intégration explicite des émotions morales dans les modèles de prise de décision économique

Par l'utilisation d'un raisonnement à la Cournot, un unique équilibre de Nash est obtenu et correspond graphiquement à l'intersection des droites représentant les fonctions de meilleure réponse des deux agents et qui est tel que des contributions positives sont possibles à l'équilibre, ce qui est en accord avec les résultats expérimentaux observés dans ce jeu, notamment dans les travaux de Fehr et Gächter (2000).

Les auteurs présentent ensuite des résultats de statique comparative et de simulation.

Tout d'abord, ils analysent le cas homogène, dans lequel les agents ont des paramètres identiques. Les résultats montrent que, pour chaque agent $i \in I$, une augmentation du paramètre de honte σ_i , ou du paramètre de culpabilité β_i , se traduit par une augmentation du niveau des contributions, ainsi que par une baisse du montant des punitions infligées. De même, une augmentation du paramètre d'altruisme α_i , du paramètre de réciprocité λ_i , ou de la norme de contribution a_i^* , conduit également à une augmentation de la coopération.

En outre, les auteurs s'intéressent aussi à la confrontation entre un agent i standard et un joueur j purement égoïste, soit un agent caractérisé par des paramètres individuels étant tous égaux à zéro et qui ne punit jamais.

Les résultats démontrent que, dans certains cas (suivant les valeurs de λ et γ), le comportement de « free-riding » de l'agent égoïste peut être dissuadé par la menace d'une sanction qui n'est pas exécutée dans les faits (à l'équilibre), tandis que dans d'autres cas, une punition est finalement mise en œuvre par le joueur i , sans inciter l'agent égoïste à coopérer. Quoi qu'il en soit, dans chacun de ces deux cas, aucun agent ne contribue à hauteur de la norme de l'agent i (l'agent égoïste n'ayant pas intériorisé de norme), ce qui est illustratif des difficultés rencontrées par la punition dans la promotion de la coopération auprès d'un agent égoïste, ainsi que de la capacité d'un tel agent à tirer les contributions à la baisse, au sein de son groupe. Néanmoins, dans le dernier cas, les simulations montrent que, si l'agent j passe du statut d'un individu purement égoïste, à celui d'un joueur plus sensible au sentiment de honte, le niveau moyen de contribution des agents augmente et le montant de la punition baisse. L'ensemble de ces résultats semble à nouveau être pertinent vis-à-vis de la réalité empirique observée dans cette version du jeu du bien public.

Plus tard, Bowles et Gintis (2005) généralisent le modèle aux cas impliquant $n > 2$ joueurs, avec $i = 1, \dots, n$ et I désignant toujours l'ensemble des joueurs.

En posant w_i comme étant la valeur de la dotation initiale de chaque agent $i \in I$, la première étape du jeu conduit chaque joueur à choisir un montant de contribution $a_i \in [0, w_i]$, l'ensemble de ces contributions rapportant un gain $f(a_1, \dots, a_n)$ à chacun des agents du groupe. La fonction f est définie de sorte à être croissante dans chacun de ses arguments et, dans le même temps, de façon à ce que $\frac{\partial f}{\partial a_i} < 1 \forall i \in I$, pour garantir le fait que le jeu soit un dilemme social. Pendant la deuxième phase, chaque joueur $i \in I$ est informé des niveaux de dotations (w_1, \dots, w_n) et de contributions (a_1, \dots, a_n) de l'ensemble des joueurs du groupe et a l'opportunité d'infliger une sanction μ_{ij} à chacun des autres membres ($j \neq i$) au coût $c_i(\mu_{ij})$, par unité de punition imposée.

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

De manière similaire au cas où $n = 2$, le gain matériel du joueur i est défini par :

$$\pi_i = w_i - a_i + f(a_1, \dots, a_n) - \sum_{j \neq i} \mu_{ji} \quad (3.3.10)$$

A nouveau, pour les mêmes raisons que celles que nous avons expliquées auparavant, le coût supporté par i pour punir un autre joueur $j \neq i$ n'est pas inclus dans l'expression de π_i .

Alors, dans le prolongement du cas où $n = 2$, la fonction d'utilité du joueur i est définie par :

$$U_i = \pi_i - \sum_{j \neq i} c_i(\mu_{ij}) + \sum_{j \neq i} \delta_{ij} \pi_j - \beta_i (a_i^* - a_i)^2 - \sigma_i (a_i^* - a_i) \sum_{j \neq i} \mu_{ji} \quad (3.3.11)$$

Le modèle général se résout au moyen de raisonnements comparables à ceux utilisés dans le cas basique, et les résultats qui sont ainsi obtenus par les auteurs sont conformes à ceux présentés précédemment. En particulier, il peut être démontré, par exemple, que lorsque la contribution de j augmente, i punit moins celui-ci (ou du moins pas plus). En outre, Bowles et Gintis (2005) démontrent que si un agent contribue moins que ce qui est suggéré par sa norme, une augmentation de la force avec laquelle celui-ci ressent la honte induit une augmentation de son niveau de contribution. Enfin, les auteurs montrent également que lorsque la norme de contribution d'un agent est augmentée, celui-ci contribue plus.

La résolution du modèle s'effectue à travers une maximisation de l'utilité de l'agent, vis-à-vis de la décision de contribution et de la décision de punition. Les auteurs présentent également quelques résultats de statique comparative pertinents vis-à-vis des résultats expérimentaux obtenus dans le jeu du bien public. Ainsi, ils montrent que pour $\lambda_i > 0$, la contribution de meilleure réponse de chacun des individus augmente avec le niveau de contribution de l'autre, *i.e.* ils constatent la mise en place d'une réciprocité positive, que l'on pourrait attribuer à l'émergence de la gratitude. Ils prouvent également qu'une augmentation de la valeur des paramètres β_i et σ_i implique (que ce soit pour l'un, l'autre ou les deux), toutes choses égales par ailleurs, une augmentation du niveau de contribution à l'équilibre pour chacun des agents.

Cette modélisation admet certaines limites, comme par exemple le fait que les solutions obtenues ne sont ni très esthétiques, ni toujours facilement présentables ou interprétables. On peut également penser au fait que le jeu se déroule en information incomplète, alors que celui-ci gagnerait encore en réalisme en intégrant de l'incertitude, en ce qui concerne la personnalité des joueurs. Ces deux écueils peuvent sans doute être expliqués par le nombre élevé de paramètres présents dans le modèle, ce qui justifie aussi le recours à des procédés de simulations.

Néanmoins, on ne peut ignorer la richesse de cette formulation en ce qui concerne la représentation de l'influence simultanée de deux émotions différentes, sur le choix stratégique des individus. En outre, on note également la pertinence de ce modèle, s'agissant de la prise en compte de l'arbitrage existant entre des motivations matérielles égoïstes et des préoccupations, d'ordre émotionnel, orientées vers autrui (qui sont liées à l'examen du sort des partenaires dans le jeu et à la présence de normes de comportement dans l'économie), qui entrent en jeu lorsque les agents ont à faire un choix dans une situation de dilemme social.

En outre, ce travail témoigne de la volonté de pousser plus loin l'analyse de l'influence des émotions sur le choix des individus dans ce type de situation. En pratique, cette volonté s'observe par l'intégration précise de diverses émotions discrètes (à savoir, entre autres, la culpabilité et la honte), ce qui permet de faire une utilisation plus ciblée des recherches en psychologie sociale et en économie expérimentale. Etant donné que les humains sont des êtres sensibles, c'est par cette recherche d'intégration, la plus complète possible, des mécanismes psychologiques influents sur le choix des agents économiques, que l'on sera capable de rendre plus réaliste la représentation de leur processus de décision et que l'on obtiendra, en conséquence, une plus grande capacité prédictive des modèles comportementaux, ainsi qu'une meilleure orientation des politiques économiques afin de favoriser la mise en place de la coopération.

3.3.2.2. Un modèle d'aversion émotionnelle à la violation de la norme

Plus récemment, les travaux de Lopez-Pérez (2008) ont proposé un modèle intégrant explicitement de nombreuses émotions morales. Le point de départ de la réflexion de l'auteur est de dire que certains comportements humains (la coopération, la générosité etc.) peuvent en partie être induits par l'adhérence à des normes sociales ou morales. Ces normes, présentes dans nos sociétés, sont internalisées par les agents à travers le processus d'éducation et modifient les choix des agents par des réponses émotionnelles.

Lopez-Pérez (2008) identifie alors explicitement l'importance des émotions morales *négatives* que sont la culpabilité et la honte (émotions auto-conscientes) d'une part, et du rôle de la colère/indignation d'autre part¹³. Conformément à ce qui se dégage des éléments revus lors des chapitres précédents, l'auteur suppose que les émotions de honte et de culpabilité interviennent à la violation d'une norme sociale ou morale que l'on a internalisée. L'émergence de ces émotions, ainsi que leur anticipation, façonnent les comportements des individus puisque, toutes choses égales par ailleurs, ceux-ci préfèrent éviter de les ressentir. Lopez-Pérez (2008) souligne, en particulier, le rôle de la colère/indignation dans l'émergence des comportements punitifs à l'encontre des transgresseurs, qui incitent (conjointement avec les émotions auto-conscientes) au maintien et au respect des normes morales ou sociales.

Les agents, dans ce modèle, sont supposés être rationnels, dans le sens où chacun d'eux cherche à maximiser ses gains, étant donné ses croyances sur la stratégie de l'autre. Néanmoins, chaque agent ne valorise pas ses gains de la même façon, l'auteur prenant alors en compte, de manière originale, la personnalité des agents. Pour lui, il existe deux catégo-

13. Emotion qui apparaît implicitement chez Bowles et Gintis (2005), à travers les paramètres liés à la présence d'une réciprocité négative

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

ries d'agents au sein de la population. Tout d'abord, on trouve des agents égoïstes, qui ne valorisent que leurs gains matériels et dont la proportion dans la population est de $(1 - \mu)$. Ensuite, il existe également des « agents de principes » (de l'anglais « principled agents », notre traduction), présents en proportion μ dans la population¹⁴ et pour lesquels d'autres considérations que le seul bien-être matériel entrent en ligne de compte. En particulier, l'utilité des agents de cette dernière catégorie est impactée par la correspondance de leur action à « l'E-norm », une norme spécifique qui est supposée être la seule en vigueur dans la société.

L'« E-norm » tire son nom des deux caractéristiques normatives qu'elle représente. Pour l'auteur, les agents sont à la fois soucieux de promouvoir l'équité entre eux, mais aussi de faire des choix économiquement *efficients*, avec une préférence pour le dernier critère. Par exemple, Lopez-Pérez (2008) rapporte une étude de Charness et Grosskopf (2001) dans laquelle des sujets, ayant reçu 600 pesetas, doivent décider d'un montant, compris 300 et 1200 pesetas, à accorder à un autre joueur. D'après les résultats de l'expérimentation, 71,4 % des agents ont choisi que 1200 pesetas soient versées à leur partenaire. Ce résultat montre bien qu'au détriment d'éventuelles considérations égalitaires, une large majorité d'agents fait un choix efficient qui conduit à l'obtention d'un gain maximal pour l'ensemble des joueurs.

Formellement, pour que la norme soit respectée, l'agent doit sélectionner les actions permettant de se trouver sur un sentier tel que l'allocation finale du jeu maximise la fonction suivante :

$$F_{\epsilon\delta} = \epsilon \cdot \sum_{i \in N} x_i - \delta(\max_{i \in N} x_i - \min_{i \in N} x_i) \quad (3.3.12)$$

Où les auteurs supposent $\epsilon = 1$ et $0 < \delta < 1$, du fait de la préférence pour le critère normatif d'efficience.

Parallèlement, l'auteur définit la fonction d'utilité relative à chaque type d'agent. Ainsi, pour chaque nœud terminal z du jeu considéré, on peut définir la fonction d'utilité d'un agent égoïste comme :

$$u_i(z) = x_i(z) \quad (3.3.13)$$

En outre, la fonction d'utilité de « l'agent de principes » intègre, bien entendu, les gains matériels de cet agent dans le jeu. Toutefois, comme celui-ci a internalisé « l'E-norm », il subit un coût psychologique lorsqu'il choisit un comportement qui ne lui est pas conforme.

Une des spécificités du modèle est que l'intensité de ce coût dépend négativement du nombre de transgresseurs dans la population. Cet aspect est, d'après l'auteur, une manière de tenir compte de l'effet de l'émotion de honte, laquelle est en lien avec le sentiment

14. Chaque agent connaît la distribution des types au sein de la population (μ est connu de tous), mais ne connaît pas spécifiquement le type de chacun des autres agents.

d'infériorité et avec la manière dont on peut comparer son comportement avec celui des autres. Grossièrement, l'idée est qu'un « agent de principes » ressent moins de honte à transgresser une norme, lorsqu'un nombre conséquent d'individus ne respecte pas non plus la règle, en raison du fait qu'il est supposé ressentir ce sentiment uniquement vis-à-vis des personnes dont le comportement est moralement meilleur que le sien.

Afin de formaliser cet aspect, l'auteur définit $R(z)$ comme l'ensemble des joueurs respectant la norme dans l'histoire aboutissant à z et indique que $r(z)$ représente le cardinal de cet ensemble. A l'aide de cette notation, il définit la fonction d'utilité de l'agent de principe comme :

$$u_i(z) = \begin{cases} x_i(z) & \text{si } i \in R(z) \\ x_i(z) - \gamma.r(z) & \text{si } i \notin R(z) ; (\gamma > 0) \end{cases}$$

Où γ représente l'intensité avec laquelle l'agent a internalisé la norme.

Lorsque l'agent ne respecte pas « l'E-norm », le terme $\gamma.r(z)$ vient diminuer son utilité et représente un coût psychologique lié au fait de ressentir des émotions désagréables. De manière assez originale, l'auteur ne modélise pas de manière distincte le rôle de la honte et de la culpabilité, mais tient plutôt compte, à travers ce terme, d'un rôle conjoint de ces deux émotions. En effet, on peut retrouver, au sein de ce terme, des caractéristiques propres à chacune d'entre-elles. D'un côté, le paramètre d'intensité de l'internalisation peut être considéré comme lié à la dimension « autocentrée » de l'évaluation que l'on fait de nos propres comportements, laquelle se rapporte plus à l'émotion de culpabilité. De l'autre côté, le paramètre représentant le nombre de transgresseurs est relatif à la dimension publique de l'auto-évaluation de nos comportements et des jugements qu'ils suscitent, ce qui est sans doute plus en lien avec le sentiment de honte.

La résolution du modèle s'effectue à l'aide du concept d'équilibre bayésien parfait et les équilibres obtenus dépendent alors des valeurs des paramètres μ et γ . L'auteur montre qu'une plus grande proportion « d'agents de principes », au sein de la population, ainsi qu'un coût psychologique plus important lors d'une déviation, implique une augmentation des comportements respectueux de la norme chez les « agents de principes » (et parfois même chez les agents opportunistes), dans des situations telles que les jeux à la cournot, les jeux de confiance ou le jeu du centipède.

Afin d'intégrer l'émotion de colère, la fonction d'utilité des agents de principes est légèrement complexifiée. En supposant un jeu à N agents :

$$u_i(z) = \begin{cases} x_i(z) & \text{si } R(z) \equiv N \\ x_i(z) - \gamma.r(z) & \text{si } i \notin R(z) ; (\gamma > 0) \\ x_i(z) - \alpha \max_{j \notin R(z)} x_j & \text{si } R(z) \subset N, i \in R(z) ; (1 \geq \alpha > 0) \end{cases} \quad (3.3.14)$$

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

Ainsi, lorsque l'on tient compte de la colère, la fonction d'utilité d'un « agent de principes » i se compose, selon les circonstances, de différents éléments. Comme dans le cas précédent, lorsque tous les agents (lui y compris) respectent la norme, celle-ci intègre seulement le gain monétaire de l'agent i . De même, lorsque l'agent i dévie, le niveau d'utilité qu'il obtient correspond à son gain monétaire dans le jeu, diminué du coût psychologique lié à la violation de la norme, ce coût dépendant toujours du nombre d'agents qui la respectent. Enfin, lorsque l'agent i respecte la norme mais qu'au moins un des autres joueurs dévie, son utilité i équivaut au gain monétaire obtenu dans le jeu, auquel est soustrait un terme désignant la colère ressentie par i du fait d'observer la violation qui est commise par autrui. Ce terme est composé du paramètre α , correspondant à l'intensité de l'agressivité de l'agent, lequel multiplie le niveau de gain du « pire » joueur déviant, soit celui qui s'en sort avec le gain le plus élevé. La colère de l'agent se concentre donc uniquement sur le partenaire qui profite le plus de la déviation qu'il commet. De plus, il convient de remarquer que l'intensité de la colère de l'agent i ne dépend pas de la nature même de la déviation particulière qui est réalisée, en termes de l'action qui est choisie, mais bien des conséquences matérielles qui en résultent. Malgré le fait que ces hypothèses puissent sembler un peu fortes, l'auteur explique qu'une telle modélisation se confronte plutôt bien aux données empiriques.

Ce terme de colère, en venant diminuer l'utilité de l'individu, façonne son choix et peut, par exemple, pousser le répondant d'un jeu de l'ultimatum, à rejeter l'offre (trop) injuste qui lui est faite. Dans un tel cas, le répondant préfère se venger et obtenir un gain matériel nul, plutôt que de profiter d'un gain positif, tout en souffrant du fait d'être injustement traité, à travers la colère qui émerge en lui. Ce phénomène peut être anticipé par le proposeur (quel que soit son type) et peut orienter celui-ci, en fonction de ses croyances sur la distribution des types dans la population, ainsi que sur le niveau de colère ou d'agressivité qu'engendre une offre injuste, vers la réalisation d'une proposition respectueuse de l'E-norm. Lopez-Pérez (2008) insiste sur le fait que, dans sa modélisation, ce n'est pas (ou pas seulement) à des fins de réduction de l'iniquité entre agents (comme dans les modèles à la Fehr et Schmidt, 1999) que les actions punitives sont réalisées, mais plutôt sous le coup de la colère ressentie par l'agent irrité. En s'inscrivant dans une telle représentation, on peut alors envisager la colère comme une émotion qui pousse l'agent à réagir violemment lorsqu'il se sent victime d'une injustice, de sorte à ce que, par la peur qu'il suscite chez les individus qui l'entourent, de telles injustices ne se (re)produisent pas.

Une limite éventuelle de ce modèle tient au minimalisme de l'hétérogénéité existant au sein la population d'agents. En effet, au lieu de considérer que le coût psychologique supporté par chaque agent puisse avoir une intensité qui soit différenciée individuellement (pouvant être nulle dans le cas d'agents égoïstes), ici, seuls deux niveaux d'intensité correspondant à deux types spécifiques d'agents ($\gamma = 0$ pour les égoïstes ou γ est une constante non nulle pour les « agents de principes ») sont envisagés. L'auteur fait donc d'emblée l'hypothèse que chaque « agent de principes » souffre avec la même intensité lors de la violation d'une norme, ce qui est une hypothèse simplificatrice assez peu réaliste (et qui pourrait être éliminée en indiquant le paramètre d'intensité de l'internalisation). Toutefois, cette manière de modéliser permet de mettre en évidence, de façon plus soutenue, le fait que certains individus ne se soucient aucunement de la norme lors de leurs choix. De plus, une telle formalisation permet également de tenir compte du fait que la pertinence d'une norme peut être liée au nombre d'individus qui font le choix de la respecter.

Une autre limite inhérente au modèle de Lopez-Pérez (2008), laquelle est commune à l'ensemble des travaux revus dans ce chapitre, tient au fait que certains auteurs considèrent que, s'agissant de la prise en compte de l'impact des normes morales et des émotions morales sur la prise de décision des agents, l'utilisation d'une théorie du choix rationnel n'est pas la plus adaptée.

3.3.3. Théorie du choix rationnel et prise en compte du rôle des normes et émotions morales

Ce point de vue est notamment développé dans le travail de Vanberg V. (2008). D'après lui, s'agissant de l'influence des règles et des normes morales sur le comportement des agents, le choix des individus ne se porte pas au niveau des *résultats* espérés pour les différentes stratégies possibles. En effet, dans l'hypothèse inverse, l'action entreprise n'est que le moyen stratégiquement employé afin d'obtenir le meilleur résultat possible, ce qui aboutit à un système de préférence sur les résultats. Selon lui, le choix se porte plutôt au niveau des *actions* possibles pour l'agent, avec la présence de préférences pour certaines actions, en soi, du fait de leur congruence avec les critères normatifs de l'individu, soit un système de préférence sur les actions. Toujours en accord avec le propos de Vanberg V. (2008), la prise en compte de l'influence des normes morales, notamment à travers le rôle des émotions morales, nécessite, par rapport au cadre classique, un changement de focus entre une théorie du choix rationnel et une théorie des comportements respectueux des règles. En particulier, l'auteur insiste sur le fait que les critères normatifs moraux des individus ne peuvent pas être représentés dans la fonction d'utilité du fait qu'il n'est pas possible de mettre sur le même plan les préférences pour certaines actions *per se* et les préférences qui sont liées au résultat d'une action, notamment en raison du problème de la commensurabilité entre préférences morales et préférences matérialistes. De plus, selon Gagné (2007), un autre problème réside dans le fait d'intégrer des émotions dans la fonction d'utilité, puisque, par exemple, les émotions ont la capacité de prendre le pas sur les processus cognitifs à l'œuvre lors d'une décision, ce qui implique que les émotions ne devraient pas seulement modifier le résultat de la maximisation de la fonction d'utilité, mais également la fonction d'utilité elle-même.

Avec la théorie des comportements respectueux des règles, les agents apprennent et déterminent, à travers plusieurs processus (biologique, culturel, individuel), des actions adaptées, sous la forme de normes ou de règles de comportement, s'appliquant dans les différents types de situations auxquels ils sont confrontés. Par la suite, de tels individus agissent en fonction de ces règles de comportement lorsqu'ils font face à un choix. Ainsi, là où un agent rationnel parfait, *i.e.* un agent possédant des informations complètes et doté de capacités de raisonnement parfaites, peut procéder à un calcul de maximisation afin de faire le choix stratégique optimal en toute situation, un agent « imparfait », au sens d'un agent qui n'est pas doté de telles capacités, peut avoir globalement intérêt à moduler son choix par le respect et l'application de certaines règles, bien que cela le conduise, en certaines occasions, à prendre une décision sous-optimale, plutôt qu'à mettre systématiquement en œuvre une procédure de maximisation, potentiellement source d'erreur. De plus, les capacités nécessaires à la mise en œuvre d'un choix rationnel de maximisation peuvent sembler, dans de nombreuses circonstances, hors de portée pour le commun des mortels. Dans cette perspective, s'il est nécessaire de comparer ces deux modèles de re-

3. Une Intégration Graduelle du Rôle des Emotions dans la Théorie Economique

présentation, la théorie des comportements respectueux des règles semble être celle qui propose le plus haut degré de réalisme.

Toutefois, Vanberg V. (2008) explique que les deux systèmes de préférence ne sont pas incompatibles, bien que mutuellement excluables. En effet, les agents peuvent passer de l'un à l'autre. En particulier, les individus peuvent préférer, au cas par cas, utiliser un système de préférence sur les résultats (à travers un processus de maximisation), plutôt qu'un système de préférence lié aux actions, notamment lorsque la perspective de gains (matériels) personnels conséquents est présente.

La théorie soutenue par Vanberg V. (2008) laisse également une place importante aux émotions, dans la mesure où celles-ci incitent au maintien des agents dans un système de préférence sur les actions les conduisant à respecter les normes morales qu'ils ont intériorisées.

Au total, malgré la pertinence de son analyse et de sa critique sur le manque de réalisme de la théorie du choix rationnel, l'approche de Vanberg V. (2008) n'est pas aussi éloignée que l'on pourrait le croire des modèles de préférences sociales que nous avons évoqués et de certaines de leurs caractéristiques. En effet, dans chaque cas, les trois propriétés suivantes sont inhérentes à la modélisation : (1) l'agent est sensible à la présence de normes morales lors de sa prise de décision, (2) le choix final de l'agent est vu comme un arbitrage entre son bien-être personnel et le respect des normes morales s'appliquant à lui et il est possible que l'incitation à privilégier la satisfaction individuelle soit suffisamment importante pour le conduire à la violation de ces normes, et, (3) à l'inverse, l'émergence des émotions (morales) joue un rôle dans la promotion des comportements socialement adaptés.

Ainsi, bien qu'une approche mathématisée, à l'aide de fonctions d'utilités enrichies, semblables à celles utilisées dans la plupart des travaux d'économie comportementale, puisse être mise en doute conceptuellement, elle représente toutefois une approximation pertinente et nécessaire, en vue d'obtenir des outils permettant des prédictions quantifiables.

De plus, en recherchant sans cesse plus de réalisme lors de l'élaboration des fonctions d'utilité, en s'appuyant notamment sur les résultats obtenus en économie expérimentale, en psychologie sociale ou dans les travaux de neuro-sciences, les modèles conçus en économie comportementale gagnent en capacité prédictive.

Enfin, la nuance qui est apportée par l'étude spécifique et précise des mécanismes psychologiques, et notamment l'analyse discrète des émotions, peut également nourrir l'inspiration des modélisateurs et ainsi être à l'origine même de la conception de nouvelles formes fonctionnelles originales, permettant de mieux représenter (et prédire) les comportements dans certaines situations particulières. Par exemple, l'étude de l'émotion de honte est venue nourrir la modélisation de Tadelis (2008), ce qui a permis à celle-ci de rendre compte de la différence de comportement qui est induite par une différence d'exposition des actions entreprises par les joueurs.

3.4. Conclusion

Cette revue a permis de mettre en évidence les progrès qui ont été réalisés par les économistes dans leur manière de représenter le processus de décision des agents économiques dans les situations d'interaction sociale. Le relâchement de l'hypothèse de la recherche de l'intérêt individuel, à travers l'intégration de mécanismes psychologiques liés à l'influence des normes (morales ou sociales), des intentions ou des attentes des agents, a conduit à la création de modèles plus réalistes et donc plus fidèles à ce qui caractérise les êtres humains et leur manière de fonctionner en société.

Les différents mécanismes psychologiques développés dans ces modèles, dits de préférences sociales, sont profondément en lien avec l'émergence et l'influence des émotions morales qui, de manière implicite ou explicite, trouvent leur place au sein des différentes formalisations présentées dans ce chapitre.

Néanmoins, il convient de réaliser qu'à mesure que les connaissances scientifiques deviennent de plus en plus précises vis-à-vis des ressorts psychologiques sous-tendant les comportements humains, l'intégration progressive des nuances qui émergent, notamment, au sein des littératures en psychologie ou en économie expérimentale, est une voie particulièrement séduisante s'agissant de conduire l'économiste vers un niveau supérieur de réalisme dans la représentation, et donc dans la prédiction, de ces comportements.

Ainsi, plus l'intégration des émotions (en particuliers des émotions morales) s'effectuera de manière explicite et précise dans les travaux de modélisation, notamment à travers la reconnaissance des particularités singulières des différentes émotions, plus ces travaux seront à même de représenter le plus fidèlement possible le fonctionnement des déterminants des choix stratégiques des agents. Par conséquent, afin de continuer à nourrir la représentation par l'économiste du processus de décisions des individus, il semble nécessaire de poursuivre l'exploration de la « boîte noire » que constitue la psychologie humaine et ses différentes composantes.

4. Le jeu du bien public : un cadre approprié à la représentation du rôle économique de la gratitude

Dans ce travail de thèse, nous proposons un modèle original simple visant à analyser l'impact des émotions morales, dans leurs diverses formes. Afin de mieux rendre compte des nombreuses caractéristiques de l'influence de la gratitude, ainsi que de celle des autres émotions morales, sur le comportement des individus, nous avons choisi d'insérer notre formalisation dans le cadre riche du jeu du bien public, à l'image de modèles comme celui de Bowles et Gintis (2005) ou de van Dijk et van Winden (1997), que nous avons précédemment cités. Ainsi, dans ce chapitre, nous réalisons d'abord une présentation détaillée du jeu du bien public, avant de passer en revue les principaux résultats empiriques ayant été dégagés expérimentalement, à propos de ce jeu et de ses différentes déclinaisons. Enfin, nous concluons en expliquant pourquoi le jeu du bien public représente une situation particulièrement attrayante dans la perspective de l'analyse formelle des émotions morales.

4.1. Le jeu du bien public

Une des questions centrales caractéristiques du dilemme social est celle de la production des biens publics purs. Par nature, ces biens sont caractérisés par la non-rivalité et la non-exclusion. Précisément, la non-rivalité suppose que la consommation d'un tel bien, par un agent, n'empêche pas un autre individu de pouvoir également le consommer, tandis que la non-exclusion implique qu'aucun agent ne peut être exclu de la consommation d'un bien public. Dans nos sociétés, on trouve de nombreux exemples concrets de ces biens tels que l'information, la qualité d'un environnement naturel (qualité de l'air, biodiversité etc.), la construction d'un pont ou encore l'occupation d'un espace public.

Lorsque l'on envisage les agents économiques comme des êtres rationnels et égoïstes, au sens où ils ne s'intéressent qu'à leur propres gains matériels, peut apparaître le problème associé au financement d'un bien public pur. En effet, une fois que celui-ci est financé, chaque agent peut en profiter comme bon lui semble. Par conséquent, il est sensé, pour tout individu, de chercher à jouir de ce bien sans avoir à en supporter un quelconque coût. L'incitation matérielle des agents à adopter un tel comportement, dit de « passager clandestin », est ainsi susceptible de menacer la production optimale des biens publics, par le marché. Par conséquent, la question du financement privé et volontaire d'un bien public correspond à une situation de dilemme social, dans le sens où il est dans l'intérêt de chaque agent de ne pas contribuer, alors même que toute contribution représente un intérêt pour la collectivité.

En théorie des jeux, le dilemme du prisonnier (Tucker, 1950), où deux joueurs doivent faire un choix entre coopérer (avec le partenaire) ou ne pas coopérer, a permis d'identifier

4. Le jeu du bien public : un cadre approprié à la représentation du rôle économique de la gratitude

la situation de dilemme social, ainsi que le comportement de « passager clandestin » qui est adopté par tout *homo oeconomicus*. Le jeu du bien public, qui correspond à une version continue du dilemme du prisonnier, permet de représenter le financement volontaire d'un bien public.

En substance, dans un jeu du bien public standard, un groupe de $n \geq 2$ participants est formé et on suppose qu'une dotation initiale est fournie à chaque individu au début du jeu, qui peut être identique, pour tous les joueurs, ou différente, d'un participant à l'autre, selon les versions du jeu. Chaque individu doit ensuite décider simultanément, et sans possibilité de communication, la somme qu'il souhaite investir pour contribuer à l'instauration (au maintien ou à l'amélioration) d'un bien public bénéficiant à l'ensemble des agents, étant donné que chacun consomme individuellement le restant de sa dotation. Dans la plupart des versions du jeu, chaque joueur reçoit un gain matériel lié à la quantité de bien public financée, qui correspond à la somme des contributions réalisées par l'ensemble des membres du groupe, multipliée par un paramètre de rendement marginal du bien public¹.

Pour correspondre aux caractéristiques d'un dilemme social, le jeu est construit de manière à ce que la stratégie dominante, pour chaque individu, soit de ne rien contribuer, alors même que les paiements collectifs sont maximisés lorsque chaque joueur contribue l'intégralité de sa dotation, ce qui correspond à un rendement marginal du bien public compris entre $\frac{1}{n}$ et 1.

Dans ces conditions et en supposant que les agents sont rationnels et égoïstes, les contributions obtenues à l'équilibre de Nash sont nulles, puisqu'il est individuellement avantageux pour les joueurs de ne rien dépenser et de compter sur la contribution des autres.

Par ailleurs, le jeu du bien public peut être joué en un coup ou de manière répétée, les participants étant alors mis dans des conditions identiques à celles de la première répétition du jeu, lors de chaque nouvelle période. Lorsque le nombre de périodes jouées est fini, un raisonnement par induction à rebours permet de déterminer que l'équilibre de Nash parfait en sous jeu, pour des individus rationnels et égoïstes, correspond au fait de jouer à chaque période l'équilibre de Nash du jeu en un coup, ce qui conduit donc à ce que finalement aucun agent ne contribue au bien public pendant toute la durée du jeu².

Face à de telles prédictions théoriques, les économistes ont cherché des moyens permettant de désinciter les agents à agir comme des « passagers clandestins », de sorte à pallier les défaillances du marché à fournir un financement privé des biens publics qui soit optimal.

1. Nous proposons une présentation formelle et détaillée du jeu du bien public et de son déroulement, selon différentes versions, dans le chapitre suivant (cf. 5.1).

2. Néanmoins, lorsque le nombre de périodes est infini, des niveaux de contributions positifs peuvent être obtenus, à l'équilibre, à travers l'adoption de stratégies visant à maximiser le bien-être intertemporel des agents (on retrouve alors des stratégies similaires à celles utilisées dans le dilemme du prisonnier, comme par exemple la stratégie du « donnant-donnant », qui veut que l'on coopère à la première période, puis que l'on joue ce qu'a joué l'autre à la partie précédente ou encore, la stratégie « rancunière », qui implique que l'on coopère à la première période, mais que l'on trahisse toujours, à partir du moment où le partenaire choisit de ne pas coopérer), qui dépend notamment de leur préférence pour le présent.

A l'origine, ceux-ci se sont alors tournés vers l'utilisation d'outils tels que la fiscalité, ou le recours à des mécanismes de subvention, modifiant le rendement individuel marginal d'un bien public, de manière à rendre la contribution (plus) attractive sur le plan économique (Villeval, 2010)³. En effet, lorsque l'on suppose que le planificateur fait face à des *homo oeconomicus*, seuls de tels outils sont en mesure d'inciter efficacement les agents à contribuer au financement d'un bien public.

4.2. Résultats expérimentaux

Le jeu du bien public a néanmoins également été étudié empiriquement, dans de nombreux travaux expérimentaux, et les résultats obtenus suggèrent que le problème de « passer clandestin » n'a pas autant d'ampleur que ce qui est classiquement prédit par la théorie des jeux (voir Ledyard, 1995 et Eber et Willinger, 2012, pour une revue). Dans la majorité des protocoles, le jeu est présenté sans qu'il y ait de contextualisation (aucune référence explicite n'est faite sur la nature du bien public auquel les joueurs peuvent contribuer, ni sur la moralité des actions etc.), bien qu'il soit précisément expliqué aux sujets quelles sont les conséquences de leurs choix stratégiques, en termes de gains monétaires. Conformément aux règles de l'économie expérimentale, les participants obtiennent une rémunération réelle (en euros) à la fin de l'expérience, qui dépend du niveau de leurs gains et qui s'effectue de façon anonyme. De cette façon, les sujets sont en présence d'une réelle incitation économique, dans la mesure où leurs choix stratégiques ont des conséquences concrètes.

Ainsi, dans une situation où aucun agent n'est censé coopérer, on observe généralement que la contribution au bien public, consenti volontairement par les sujets, se situe entre 40 et 60 % du niveau de contribution socialement optimal (qui correspond à 100 % de leur dotation initiale) lorsque le jeu est joué en un coup ou encore lorsque l'on se focalise uniquement sur la première période du jeu (Chaudhuri, 2011)⁴. Ainsi, d'après Ferrari, Mekni, Petit et Rouillon (2010), la capacité du groupe à régler le problème de l'action collective est, au départ, relativement forte.

Toutefois, lorsque le jeu est répété, les données montrent que les niveaux de contribution ont tendance à décliner continûment de période en période, à mesure que de plus en plus de joueurs adoptent des comportements de « passer clandestin ».

En outre, lorsque le jeu est répété un nombre fini de fois (le nombre de périodes, allant généralement de 10 à 25, étant la plupart du temps connu des joueurs), on observe un comportement dit de « fin de jeu » qui consiste à diminuer nettement son niveau de contribution au bien public, à l'approche de la dernière période du jeu, lequel se situe *in fine* autour des 10 % de la dotation initiale (Ledyard, 1995). Ferrari *et al.* (2010) indiquent que ce phénomène pourrait être dû au fait qu'il n'est plus envisageable, pour les joueurs coopérant dans le but d'inciter les autres à faire de même, de continuer une telle stratégie lors de la fin du jeu, ce qui provoque une chute drastique des contributions lors

3. On peut également penser à des « designs » de jeu impliquant que la somme des contributions des agents doit dépasser un certain seuil pour que ceux-ci puissent profiter des bienfaits du bien public. Toutefois, même dans ce type de version du jeu, la stratégie dominante reste de ne pas contribuer.

4. Toutefois, il existe de larges variations au niveau des contributions individuelles, qui s'échelonnent de 0 % à 100 % de la dotation initiale fournie aux agents.

de la dernière période. De manière complémentaire, du fait de l'arrêt total de l'expérience (et donc de l'interaction), à la fin de la dernière période, les comportements de « fin de jeu » pourraient également provenir du fait que les individus considèrent l'effort contributif inutile, dans la mesure où celui-ci était précédemment consenti dans le but d'obtenir l'acceptation de la part de leurs partenaires, de maintenir l'image de quelqu'un de moral, ou encore de participer à développer la cohésion du groupe (lorsque celui-ci reste inchangé au cours des différentes périodes).

L'ensemble de ces résultats expérimentaux remet ainsi en question l'approche classique, en théorie des jeux, du processus de décision des agents économiques et, par la même, conduit les chercheurs à creuser de nouvelles pistes dans la recherche d'outils de politiques économiques efficaces, dans la promotion de la coopération entre les agents, lorsqu'ils doivent contribuer au financement volontaire d'un bien public. Pour ce faire, il est nécessaire de s'intéresser aux comportements des individus et notamment aux déterminants de leur choix, ou non, d'adopter volontairement une conduite coopérative (Villevall, 2010). Ainsi, de nombreux travaux empiriques ont recouru à de nouveaux protocoles présentant des aménagements ou des extensions du jeu du bien public original (ou standard), afin de découvrir ce qui pèse dans le choix des individus, ainsi que dans le but de déterminer les « designs » de jeux qui permettent de promouvoir la coopération dans le jeu du bien public.

Durant la suite de cette section, nous allons présenter brièvement les principaux leviers comportementaux permettant de favoriser la coopération dans le jeu du bien public, qui ont pu être mis en évidence expérimentalement, à travers l'utilisation de protocoles originaux.

4.2.1. Différentes déclinaisons du jeu

Dans notre présentation du jeu standard et des résultats empiriques généraux qui en sont ressortis, nous avons déjà mis en lumière la pluralité des protocoles expérimentaux utilisés, vis-à-vis de la question de la temporalité. Dans certaines expériences, le jeu du bien public est joué seulement en un coup, tandis que dans d'autres (la majorité), le jeu est répété pendant un nombre (le plus souvent connu des sujets) fini de périodes. Cette distinction est essentielle pour se rendre compte de l'évolution des comportements des agents au cours du temps.

Ainsi, une première manière de comprendre les comportements de contribution « anormalements » coopératifs, qui sont observés dans le jeu du bien public en un coup, a été de considérer qu'il s'agit d'une confusion commise par certains agents. Ainsi, selon cette explication, lorsque le jeu est répété, certains participants commencent par se « tromper », lors de la première période, en contribuant un montant positif et, à mesure que le jeu se poursuit, ceux-ci apprennent de leurs erreurs, ce qui explique la tendance à la diminution du niveau de contribution rapprochant, de période en période, les comportements des agents de l'équilibre de Nash.

Afin de tester la validité de cette hypothèse, Andreoni (1988) propose un protocole original consistant à rajouter de nouvelles périodes de jeu, par surprise, après que le nombre d'interaction initialement indiqué aux sujets ne soit terminé. Les résultats montrent que le niveau des contributions a alors considérablement augmenté après la reprise du jeu, ce

qui n'est pas cohérent avec l'hypothèse d'un apprentissage des individus, quant au fait de déterminer l'équilibre de Nash du jeu du bien public. Andreoni (1995) a ainsi avancé que le déclin de la coopération serait plutôt dû à la frustration d'avoir tenté en vain de susciter un comportement bienveillant, à travers la gentillesse, plutôt qu'à l'apprentissage des incitations à se comporter en « passager clandestin ».

4.2.2. La composition dynamique des groupes

Une autre hypothèse permet d'expliquer de forts niveaux de contributions au départ du jeu, ainsi que la dégradation des contributions au cours du temps. En effet, contribuer fortement au début du jeu et diminuer progressivement le niveau de coopération, à mesure que l'on se rapproche de la fin, peut être envisagé comme une stratégie rationnelle pour des agents en condition d'information incomplète, à propos de la rationalité des autres membres de leur groupe. Ainsi, pour un joueur qui pense qu'un partenaire peut agir de manière irrationnelle (en coopérant), le fait d'adopter un comportement de « passager clandestin » revient à lui donner une indication de la marche à suivre pour profiter de l'action des autres, ce qui précipite l'apparition de l'équilibre de Nash, lequel est sous optimal. Cependant, lorsque la fin du jeu approche, il y a de moins en moins d'incitation à manipuler l'information détenue par les autres et les niveaux de contributions tendent à diminuer et à se rapprocher de zéro.

Afin de tester la validité de cette dernière hypothèse, une autre distinction de protocole est également présente dans le travail d'Andreoni (1988) et concerne la composition des groupes de participants, au cours du temps. Pour une partie des sujets, le groupe au sein duquel ils jouent reste le même tout au long du jeu et on qualifie de « Partner » un tel traitement. Pour l'autre partie, les groupes dans lequel les joueurs sont répartis sont reformés à chaque répétition de l'interaction. Le traitement auquel appartiennent ces derniers participants est désigné par l'appellation « stranger ».

Les données de l'expérience menée par Andreoni (1988) montrent alors qu'en moyenne, le niveau de contribution observé est supérieur pour les individus du traitement « stranger », comparativement à celui des participants du traitement « partner ». Ce dernier résultat est donc en contradiction avec l'hypothèse examinée, puisque celle-ci nécessite que les niveaux de contributions soient plus élevés dans le cas du traitement « partner », puisqu'il n'y a que dans ce traitement que les agents peuvent manipuler les croyances des autres d'une période à l'autre (les groupes restant les mêmes).

Une explication alternative à la présence de contributions positives, au départ du jeu du bien public, est alors de considérer que certains agents ont des préférences sociales (Fischbacher et Gächter, 2010), autrement dit des préférences pour ce qui concerne autrui (« other-regarding preferences »), et qu'ils possèdent par exemple un « goût de donner » leur permettant de faire preuve d'altruisme (Andreoni, 1989).

Plus généralement, comme nous avons pu le voir lors des précédents chapitres, une large partie des individus a des préférences sociales impliquant qu'ils sont disposés à coopérer, uniquement lorsque les autres le sont également, on appelle ces agents des coopérateurs conditionnels (Fischbacher, Gächter et Fehr, 2001). Ces joueurs peuvent alors contribuer de forts montants au financement du bien public, au départ du jeu, de sorte à encourager

4. Le jeu du bien public : un cadre approprié à la représentation du rôle économique de la gratitude

les autres coopérateurs conditionnels (potentiels) de leur groupe, à consentir des niveaux conséquents de contribution en retour. La présence d'agents adoptant des comportements de coopérateurs conditionnels, dans le jeu du bien public, a ainsi été démontrée lors d'expériences en laboratoire et sur le terrain (Frey et Meier, 2004 ; Croson, 2007 ; voir Gächter, 2007, pour une revue).

Par ailleurs, de telles préférences permettent également d'expliquer comment d'autres données expérimentales ont montré que la coopération est plus soutenue, notamment au cours du temps, dans des groupes de la condition « partner », que dans ceux de la condition « stranger » (Keser et van Winden, 2000). En effet, il n'est pas stratégique, pour des coopérateurs conditionnels, de susciter la coopération réciproque des individus de leur groupe, dans la condition « stranger », puisque la composition du groupe change à chaque période.

Néanmoins l'expérience conduite par Fischbacher et Gächter (2010) questionne le rôle des préférences pour la coopération conditionnelle. En effet, leur protocole présente une première série de tests durant laquelle il est demandé aux individus de déterminer le niveau de contribution qu'ils choisiraient pour chacun des niveaux de contribution moyens possibles de la part des autres membres de leur groupe (allant de 0 à 20, la valeur de la dotation initiale individuelle), afin d'éliciter leurs préférences. Ensuite, ces participants jouent à un jeu du bien public répété (traitement « stranger ») et, à chaque période, avant de prendre leur décision de contribution, les sujets doivent donner leurs croyances sur les contributions qui seront réalisées par leurs partenaires.

Les résultats obtenus font état de plusieurs phénomènes. Premièrement, les résultats de la première série de test montrent qu'il existe de l'hétérogénéité, au sein de la population de participants, en ce qui concerne les préférences pour la coopération. Certains ne coopèrent jamais, d'autres coopèrent tout le temps, tandis qu'une autre catégorie d'agent coopère de manière conditionnelle au comportement adopté par leurs partenaires. Ce résultat indique une disparité des préférences au sein des populations d'agents et, en ce sens, il fait écho au travail expérimental d'Offerman, Sonnemans et Schram (1996), qui témoigne de la diversité des participants au jeu du bien public, en termes « *d'orientation sociale* », une typologie regroupant plusieurs catégories d'attitude adoptée globalement par les individus, lorsqu'ils font face à une personne anonyme ⁵.

Deuxièmement, les contributions, ainsi que les croyances sur les niveaux de contribution déclinent au cours du temps et, à chaque période les croyances sont, en moyenne, toujours supérieures aux niveaux de contributions effectifs.

Enfin, les auteurs montrent que c'est la baisse des contributions qui est à l'origine de la chute des croyances et non l'inverse. Fischbacher et Gächter (2010) suggèrent alors que les individus ont un goût pour le fait de contribuer moins que les autres au bien public, si bien qu'à chaque période, les contributions effectives sont inférieures aux croyances des individus, ce qui tend à diminuer le niveau des croyances et à conduire les agents à réduire à nouveau leur niveau de contribution lors de la période suivante et ainsi de suite.

5. Nous donnons de plus amples détails sur ce point plus tard dans la thèse (cf. 5.2.1.1.2).

Toutefois, la présence dans la population d'agents individualistes, au sens où ils ne s'intéressent qu'à leur propre bien-être matériel et adoptent (le plus souvent) un comportement de « passager clandestin », permet d'expliquer la faiblesse du niveau moyen des contributions relativement au niveau moyen des croyances, si l'on considère que ces agents (individualistes) sont en mesure d'anticiper des comportements coopératifs de la part d'autrui et que les coopérateurs conditionnels sous-estiment les comportements égoïstes ⁶.

En outre, étant donné que, dans l'expérience de Fischbacher et Gächter (2010), les groupes sont modifiés à chaque période, l'incitation à contribuer dans le but de susciter la coopération d'autrui est moindre, ce qui peut expliquer la tendance des agents à réaliser une contribution inférieure à leur croyance sur le comportement de leurs partenaires. Plus généralement, le jeu du bien public est constitué de telle manière qu'il est économiquement plus intéressant de ne pas coopérer. Ainsi, même si les coopérateurs conditionnels savent ce qu'il convient de contribuer, conformément à leurs croyances sur le comportement des autres (voire sur la base de normes de comportement qu'ils ont pu intériorisées), ils peuvent être tentés de s'écarter, à la marge, de ce qu'ils jugent être juste, en raison de l'incitation économique à dévier qui est inhérente à ce contexte particulier.

La préférence pour le fait de contribuer moins que les autres, suggérée par Fischbacher et Gächter (2010), pourrait alors se résumer à une certaine forme d'indulgence ou de manque de discipline des agents, les conduisant à s'autoriser un peu plus d'égoïsme qu'il ne faudrait.

Pour autant, les individus ne font pas forcément preuve de la même indulgence à l'encontre de leurs partenaires et, d'une période à l'autre, le fait que les niveaux de contribution obtenus ne répondent pas aux croyances formulées (et donc aux attentes), notamment en raison de la présence d'agents individualistes, est susceptible de faire émerger les émotions négatives de colère ou de déception chez les joueurs les plus coopératifs, et ainsi d'expliquer le déclin des niveaux moyens de croyances et de contributions, au cours du temps.

Cette dernière interprétation est cohérente avec l'évolution dynamique des liens sociaux, conceptualisée par van Dijk et van Winden (1997) (cf. 3.2.3), et testée dans plusieurs travaux expérimentaux s'inscrivant dans le contexte du jeu du bien public. Par exemple, dans l'étude de van Dijk, Sonnemans et van Winden (2002), « *l'orientation sociale* » des sujets (décrivant, pour rappel, l'attitude d'un individu face à une personne *anonyme*) est déterminée lors d'une première étape, avant que ceux-ci ne participent à plusieurs répétitions d'un jeu du bien public à deux joueurs, en condition « partner ». Une fois le jeu terminé, une dernière étape du protocole, visant à évaluer l'attitude des participants face à leur *partenaire* dans le jeu est conduite. Le lien social qui s'est créé au cours de l'interaction est alors mesuré, pour chaque individu, comme la différence existant entre l'évaluation de son « *orientation sociale* » et celle de son attitude face à chacun de ses partenaires, à la fin du jeu.

6. Offerman *et al.* (1996) montrent que les agents ont tendance à surestimer leur croyance à propos du comportement de contribution d'autrui.

4. Le jeu du bien public : un cadre approprié à la représentation du rôle économique de la gratitude

En premier lieu, les résultats obtenus mettent en évidence que les individus réagissent au comportement de leur partenaire, dans le jeu du bien public, puisque les données montrent qu'une augmentation (diminution) du niveau de contribution de la part d'un agent est souvent synonyme d'une augmentation (diminution) de la contribution de son partenaire à la période suivante, ce qui semble confirmer l'existence de comportements de coopération conditionnelle, ainsi que d'une certaine réciprocité entre les protagonistes. Les auteurs considèrent également que ce résultat est une preuve de l'évolution du lien social entre les partenaires et que le développement de celui-ci est en partie responsable du changement d'attitude qui s'observe d'une période à l'autre. Schématiquement, une augmentation (diminution) de la contribution d'un agent favorise le développement positif (négatif) du lien social existant entre lui et son partenaire, sous l'effet des émotions positives (négatives), comme la gratitude (la colère), ressenties par ce dernier, ce qui le motive alors à agir en retour et à augmenter (diminuer) sa contribution à la période suivante. De fait, il semble que l'évolution du lien social existant entre deux agents soit un déterminant des comportements de coopération conditionnelle.

Dans les faits, l'évolution dynamique du lien social caractérisant un couple d'agents, observée à travers le changement « *d'attitude sociale* » d'un agent envers son partenaire (qui est d'abord inconnu, avant d'être identifiable sur la base des interactions que l'agent a partagées avec lui), montre qu'un lien négatif est créé chez un individu lorsque son partenaire adopte un comportement de « passager clandestin » (notamment lors de la dernière période), alors que, lorsque l'interaction se poursuit positivement au cours du temps, un lien social positif est créé au sein du binôme. Pour les agents ayant réalisé une interaction réussie, les réponses obtenues dans des questionnaires de débriefing font même état d'un désir de jouer ensemble à nouveau.

Des résultats similaires ont été dégagés par Sonnemans, van Dijk et van Winden (2006), en utilisant un protocole très proche de celui de van Dijk *et al.* (2002), à ceci près qu'ils analysent la dynamique des liens sociaux au sein de *groupes* d'individus (4 joueurs) plutôt que vis-à-vis de *couples* d'agents. Au sein des groupes, le fait qu'une interaction prolongée ait lieu entre les agents conduit également au développement de liens sociaux entre eux, dont la valence (positive ou négative) dépend de la réussite ou non de l'interaction, en termes de niveaux de coopération observés dans chaque groupe, au cours du temps.

Par ailleurs, le développement de ces liens est tel que les individus semblent être capables de discriminer entre les différents membres de leur groupe (les joueurs étant identifiés, dans chaque groupe, par les lettres A, B, C et D), au sens où l'évolution des liens sociaux établis avec des partenaires différents, varie en fonction des comportements individuels. Ce résultat est cohérent avec les conclusions de Rege et Telle (2004), dont les données expérimentales indiquent que l'*identification* des sujets, ainsi que l'*observabilité* de leurs comportements augmentent significativement le niveau de coopération dans le jeu du bien public, par rapport à un protocole où les joueurs restent anonymes et ne sont pas informés du choix de leur partenaires. L'*identification* des agents, ainsi que l'*observabilité* des comportements pourraient donc être deux conditions nécessaires à l'évolution des liens sociaux dans les groupes participant, « en partner », au jeu du bien public, probablement à travers l'influence des émotions auto-conscientes (culpabilité, honte et fierté), susceptibles d'émerger chez des agents qui savent que leurs comportements leur sont attribuables et qu'ils peuvent faire l'objet du jugement des autres. Ceci pourrait, par conséquent, expli-

quer pourquoi, dans l'expérience d'Andreoni (1988), les niveaux de coopération observés dans les groupes « partner », où les sujets n'étaient ni identifiables, ni informés du comportement contributif des autres, se sont avérés moins élevés que ceux obtenus dans les groupes de la condition « stranger ».

En outre, Sonnemans *et al.* (2006) expliquent que la majorité des liens développés sont asymétriques et, d'après les auteurs, cette asymétrie proviendrait de différences interindividuelles concernant « *l'orientation sociale* » des agents, ainsi que de leurs réactions affectives. Selon nous, une telle différence d'expérience émotionnelle, au sens d'une valorisation affective, accordée à une interaction prolongée avec de mêmes partenaires, qui varie en fonction de l'individu, au sein d'un même groupe, est probablement également lié à des disparités au niveau de l'éducation (morale) reçue par les agents, notamment en termes de comportements normatifs intériorisés.

En résumé, le recours aux différentes conditions « partner » et « stranger » a permis de mettre en évidence une catégorie de joueur centrale dans la promotion de la coopération, dans le cadre du jeu du bien public, les coopérateurs conditionnels. Par ailleurs, les données de l'étude de Fischbacher, Gächter et Fehr (2001) indiquent que près de 50 % des sujets de leur expérience sont des coopérateurs conditionnels, tandis que 30 % des participants agissent de manière totalement égoïste. L'existence de ces individus permet d'expliquer nombre de phénomènes observés dans le jeu du bien public, tels que la disparité des comportements de contribution au sein des groupes, le déclin du niveau des contributions, au cours du temps, les coopérateurs conditionnels étant affectés par les comportements de « passager clandestin », ainsi que la relance de la coopération lors du redémarrage du jeu.

Sur la base des éléments que nous venons de revoir, nous pouvons être un peu plus précis. Il semble que les coopérateurs conditionnels soient sensibles à l'adoption de comportements de « passager clandestin », par d'autres joueurs, en ce sens que ces actions représentent une déception, face à leurs attentes. De plus, la colère émergeant, en eux, du fait de la non conformité entre leurs attentes et les actions effectivement entreprises par certains joueurs, contribue à développer un lien social négatif entre eux et leurs partenaires les moins coopératifs. Chez certains coopérateurs conditionnels, et d'autant plus lorsque le jeu est pratiqué en « partner », cette évolution négative tend à entraîner des comportements de réciprocité négative, qui se traduisent par une diminution de leurs niveaux de contributions, causant alors la chute des contributions moyennes. Cependant, si le jeu est interrompu, et que les groupes (soit le contexte de jeu) sont modifiés, ces agents semblent être en mesure de retrouver, au moins en partie, leur lien social affectif initial, vis-à-vis d'individus inconnus (« *l'orientation sociale* »), ce qui peut expliquer le redémarrage de la coopération.

Parallèlement, il est également possible que les coopérateurs conditionnels puissent être sensibles au fait d'observer que d'autres joueurs répondent à leurs attentes en agissant de manière coopérative. De plus, les émotions positives, comme la gratitude, émergeant dans un tel contexte, contribuent à développer un lien social positif entre les agents participant à l'effort coopératif et, finalement, le développement de ces liens est susceptible de soutenir des niveaux relativement élevés de contribution, à travers la mise en place de comportements de réciprocité positive.

Toujours d'après Villeval (2010), la conditionnalité des comportements coopératifs, chez certains individus, est cruciale car elle offre des perspectives pour limiter l'expression des comportements de « passager clandestin ». En ce sens, la méthode expérimentale a ainsi permis de mettre en évidence des mécanismes et des transformations institutionnelles permettant de promouvoir la coopération dans le jeu du bien public et qui impliquent des protocoles intégrant des étapes supplémentaires dans le déroulement du jeu.

4.3. L'ajout d'une phase de punition et/ou de récompense

Ainsi, il existe des versions du jeu dans lesquelles, à chaque période, la phase de contribution est complétée par une seconde phase. Au cours de celle-ci, les agents ont alors la possibilité de punir ou de récompenser (parfois les deux) leurs partenaires dans le jeu, sur la base d'une connaissance commune des décisions de contributions effectuées au cours de la première étape. La plupart du temps, il s'agit de punitions ou de récompenses matérielles, qui affectent alors le gain des agents dans le jeu. L'option de punir ou de récompenser étant, dans la majorité des cas, une action coûteuse pour celui qui l'emploie, le modèle classique en théorie des jeux suppose que des agents égoïstes et rationnels n'utilisent aucune de ces deux possibilités. En conséquence, un raisonnement par induction à rebours permet de conclure que l'ajout d'une seconde phase au jeu n'a aucune influence sur le comportement de contribution des agents. Finalement, l'équilibre de Nash du modèle classique est tel qu'à chaque période, la contribution au bien public est nulle et aucun agent ne punit ni ne récompense ses partenaires.

Toutefois, la mise en évidence des coopérateurs conditionnels suggèrent que nombre d'agents sont capables de réagir au comportement de contribution d'autrui, notamment à travers l'influence de certaines émotions. De plus, dans le jeu du bien public standard, lorsque par exemple, un coopérateur conditionnel est irrité par le comportement de « passager clandestin » d'un partenaire, la seule réaction qui lui est possible afin d'exprimer son mécontentement est de diminuer son propre niveau de contribution. Or, en ajoutant au jeu une phase de punition, un tel joueur a alors à sa disposition un outil lui permettant de communiquer sa désapprobation et d'obtenir « justice » en infligeant une sanction à celui qui adopte un comportement abusif. Parallèlement, le fait de pouvoir récompenser un partenaire particulier, offre à tout agent un canal direct de communication afin de lui exprimer sa gratitude.

Dans cette perspective, le recours effectif à l'utilisation de ces mécanismes incitatifs est nettement plus attendu et, dans la littérature, on trouve de nombreuses expériences ayant démontré leur usage et leur efficacité dans la promotion de la coopération.

La plus connue d'entre elles est probablement celle de Fehr et Gächter (2000), que nous avons déjà présenté auparavant (cf. 1.3.1.1.1). Pour rappel, dans cette expérience, des groupes de participants jouent à vingt répétitions du jeu du bien public. Lors des dix premières périodes, le jeu s'effectue sans possibilité de sanction alors que, durant la deuxième moitié de l'expérience, il se joue avec l'option de punition. Les résultats observés par Fehr et Gächter (2000) démontrent alors que les niveaux de contribution sont significativement

plus élevés lorsque les joueurs ont la possibilité d'infliger une sanction matérielle à leurs partenaires, mais également que cette option est effectivement utilisée, sous le coup de la colère, à l'encontre de ceux qui adoptent malgré tout un comportement de « passager clandestin ».

En outre, comme nous l'avons déjà mentionné, l'efficacité de ce mécanisme incitatif s'étend même aux situations où les punitions à disposition des agents sont immatérielles (Masclat *et al.*, 2003 ; Carpenter *et al.*, 2004) ou lorsque la possibilité de sanctionner est offerte à des individus tiers, appartenant à un autre groupe de participants (Carpenter et Matthews, 2010).

Ces différents résultats montrent qu'une large partie des agents est prêt à supporter un coût matériel afin d'exprimer leurs émotions et leur mécontentement, mais également que de nombreux individus sont sensibles aux jugements qui sont faits à propos de leurs actions. Ces derniers préfèrent alors augmenter leur niveau de contribution plutôt que de risquer subir une sanction, que celle-ci soit matérielle ou seulement psychologique. Ainsi, on peut envisager qu'une proportion conséquente des 30 % de sujets de l'expérience de Fischbacher *et al.* (2001), ayant joué de manière totalement égoïste dans un jeu du bien public standard, n'aurait pas agi de la sorte s'ils avaient été placés dans un contexte où les décisions sont soumises au jugement et aux réactions d'autrui.

Par conséquent, cette modification du jeu semble permettre un renversement du rapport d'influence existant entre les agents qui souhaitent promouvoir la coopération et ceux qui souhaitent profiter de leurs partenaires les plus généreux (tout en étant sensibles aux jugements qui sont portés sur leurs actions). En effet, dans le jeu du bien public standard, les coopérateurs conditionnels sont à la merci des comportements de « passager clandestin » et ont tendance à converger vers ceux-ci, du fait de la réciprocité négative. A l'inverse, lorsque la punition est possible, les coopérateurs conditionnels peuvent réagir en jetant l'opprobre sur les agents les moins coopératifs, qui, en grande majorité, adaptent leur comportement face à ce nouveau pouvoir. En ce sens, la sanction est primordiale dans la promotion de la coopération, car elle permet de favoriser l'application et le maintien, dans les groupes et la société, de normes sociales (ou morales) telles que la coopération conditionnelle (Fehr et Fischbacher, 2004b). Par exemple, les résultats expérimentaux de Reuben et Riedl (2013) montrent que la punition est capable d'orienter des groupes d'agents, possédant des caractéristiques variées (en ce qui concerne leurs dotations initiales par exemple), vers différentes normes de coopération (en fonction de l'hétérogénéité établie par le protocole), alors que les niveaux de contributions convergent vers des comportements de « passager clandestin » lorsque la sanction n'est pas disponible.

Toutefois, il semble qu'une faible proportion d'individus soient indifférents à ce mécanisme incitatif et se comportent toujours en agents égoïstes et rationnels, qu'importe le coût que représentent les sanctions qu'ils subissent en agissant ainsi.

Par ailleurs, le fait que la punition soit coûteuse implique que la décision d'appliquer une mesure punitive contre un individu non coopératif possède les caractéristiques d'un dilemme social et ainsi que des « passagers clandestins de second ordre » peuvent apparaître. En effet, l'effectivité des sanctions est nécessaire pour que la menace pesant sur les agents individualistes soit crédible. Pour autant, certains coopérateurs peuvent préférer

éviter le coût de la punition et se reposer sur les sanctions mises en œuvre par les autres. Lorsque c'est le cas, l'efficacité de la punition dans la promotion de la coopération est grandement remise en question.

En outre, les émotions négatives impliquées dans le processus de punition comportent, comme nous l'avons mentionné précédemment (cf. 1.3.1.3), des effets non désirés. Par exemple, le fait d'autoriser la sanction entre les partenaires d'un groupe induit également le risque de voir survenir des punitions anti-sociales ou injustes, dirigés à l'encontre de certains coopérateurs, à des fins de vengeance par exemple (Fehr et Gächter, 2000), ce qui peut aboutir à une escalade des représailles au sein d'un groupe. Enfin, la punition entraîne une perte de ressources (et/ou d'utilité), à la fois chez celui qui la met en œuvre, comme chez celui qui se la voit infliger, ce qui met en doute la capacité de cette incitation à agir pour le bien-être social. Ainsi, on constate que l'option de punition permet d'obtenir des niveaux de contribution qui sont effectivement plus élevés que dans le jeu standard, bien que cela se fasse globalement au détriment de la qualité des relations entre les agents et que la sanction accuse une perte économique irrécouvrable lorsqu'elle est employée.

Parallèlement, on trouve dans la littérature des versions du jeu présentant une alternative au mécanisme incitatif de la sanction, dans lesquelles est offerte aux agents la possibilité de récompenser leurs partenaires. Il est fréquemment supposé que la punition représente une incitation beaucoup plus efficace que la récompense, en ce sens que les « passagers clandestins » n'encourent aucun risque de représailles, pour leurs mauvais comportements, lorsque la punition n'est pas disponible. De fait, une large partie des travaux se proposant d'étudier l'influence de l'ajout d'une phase de récompense au jeu du bien public, examine également celle de l'option de punition, dans le but de comparer l'efficacité de ces deux mécanismes incitatifs, mais aussi d'observer l'effet obtenu lorsque ces deux options sont combinées (Sefton, Shupp et Walker, 2007 ; Rand, Dreber, Ellingsen, Fudenberg et Nowak, 2009 ; Sutter, Haigner et Kocher, 2010 ; Choi et Ahn, 2013).

Ainsi, d'après l'étude de Sefton *et al.* (2007), le fait d'ajouter à un jeu du bien public répété, joué en « partner », une phase offrant la possibilité aux individus de récompenser leurs autres membres de leur groupe, a un impact significatif (en moyenne) sur les comportements de contribution et sur les niveaux de gain obtenus par les sujets, au même titre que le fait d'offrir aux agents la possibilité de punir leurs partenaires. Toutefois, les auteurs notent certaines différences de fonctionnement pour ces deux mécanismes. En particulier, ils soulignent que les comportements de « fin de jeu » sont plus précoces et plus prononcés dans la condition « récompense » que dans la condition « punition ». Par ailleurs, ils observent également que le montant de la sanction infligée à un agent est fortement corrélé avec la déviation (négative) de sa contribution vis-à-vis de la contribution moyenne de son groupe, tandis que le niveau de récompense reçu n'est que faiblement corrélé à la déviation (positive) de la contribution de l'individu, par rapport à la moyenne des contributions du groupe. En conséquence, malgré le fait que la condition « récompense » permette d'obtenir de meilleurs rendements, en terme de gains matériels, que la condition « punition » (les sanctions représentant une perte économique), les auteurs considèrent que sanctionner est un outil plus intuitif et plus fiable dans le but promouvoir la coopération dans le jeu du bien public, notamment en ce qui concerne les derniers instants de l'interaction.

Les travaux de Choi et Ahn (2013) remettent également en cause la capacité de l'option de récompense à motiver la coopération, en suggérant que les ressorts de son utilisation sont différents de ceux de la punition et n'ont pas la même portée morale. En effet, leurs résultats expérimentaux montrent que, dans un jeu du bien public répété, joué un nombre fini de fois, les niveaux de coopération obtenus dans les traitements « récompense » et « punition » sont semblables⁷ lorsque les groupes de participants ne changent pas d'une période à l'autre (« partner »), alors qu'à l'inverse, lorsque ceux-ci sont reformés avant chaque répétition (« stranger »), la punition est bien plus efficace que la récompense. L'explication avancée par les auteurs, pour expliquer cette différence de comportement entre les conditions « partner » et « stranger », est alors de considérer que seule la punition est un comportement intrinsèquement motivé, ayant pour origine la violation d'une norme sociale, tandis que l'acte de récompense est effectué à des fins plus stratégiques, de sorte à bénéficier de contributions plus élevées dans le futur.

Au total, l'efficacité de la récompense dans la promotion de la coopération semble donc conditionné au fait que les individus interagissent ensemble, pendant toute la durée de l'interaction, ce qui n'est pas le cas pour ce qui est de l'influence de la punition. S'agissant de la punition, que l'on joue de manière répétée ou non, que le groupe dans lequel on joue change ou non à chaque période, un comportement de « passager clandestin », est toujours susceptible de déclencher le sentiment de colère avec la même intensité chez les coopérateurs, conduisant à une même mise en application des sanctions, ce qui explique la capacité d'adaptation de ce mécanisme incitatif à la pluralité des contextes. Parallèlement, la volonté de récompenser trouve, au moins en partie, son origine dans l'émergence du sentiment de gratitude. Néanmoins, les comportements suscitant une émergence intense de cette émotion peuvent être rares (peu de comportements extraordinairement altruistes) ou encore ne pas déclencher la gratitude chez ceux pour qui elle a l'influence la plus importante. En effet, les agents individualistes sont probablement moins susceptibles d'agir sous l'effet de la gratitude, voire d'être en mesure de la ressentir, en comparaison avec des agents pour qui le fait de contribuer est important. Cependant, ce sont bien les agents individualistes qui bénéficient, à moindre coût, des contributions élevées de leurs partenaires les plus coopératifs. En parallèle, ceux qui contribuent et valorisent le plus la coopération, bien qu'ils soient potentiellement sujets à ressentir la gratitude plus intensément, sont amenés à ne la ressentir que faiblement, du fait du moindre niveau de contribution affiché par leurs partenaires. Pour prendre sa pleine mesure, l'impact crucial de la gratitude et de la récompense sur les comportements de contribution, pourrait nécessiter que l'interaction s'inscrive dans une perspective dynamique, impliquant une évolution intertemporelle des relations interpersonnelles.

Ainsi, là où le fait de punir représente un mécanisme de défense invariable contre les comportements abusifs, au sens ils vont à l'encontre d'une certaine idée de justice dans le partage de l'effort coopératif, le fait de récompenser un partenaire semble plutôt représenter un acte visant à favoriser la création d'une relation (d'un lien social) solide entre des partenaires qui partagent (ou tout du moins affichent) des valeurs communes, à travers une communication approuvative. En ce sens, l'interprétation de Choi *et al.* (2013), voulant que la récompense représente seulement un investissement « stratégique », au sens de matériellement intéressé, néglige la dimension affective du processus conduisant à sa

7. Sauf en ce qui concerne les comportements de « fin de jeu », qui sont à nouveau plus accentués dans la condition « récompense ».

réalisation. La construction d'une relation entre agents nécessite une interaction prolongée et, en ce sens, la condition « stranger » n'offre pas la durée de temps suffisante pour qu'un lien social fort se crée entre les partenaires, ce qui perturbe le fonctionnement du mécanisme incitatif de récompense et explique pourquoi il perd en efficacité dans ces circonstances. Une telle interprétation permet d'expliquer pourquoi, dans le cadre d'un jeu du bien public joué en coup, Walker et Halloran (2004) observent que l'option de récompense n'influe pas significativement sur les niveaux de contribution, en comparaison avec ce qui est obtenu pour des groupes « contrôles » (où elle est absente), alors même qu'elle est largement utilisée par les participants lorsqu'elle est disponible. Il convient cependant de noter que les auteurs trouvent un résultat similaire en ce qui concerne la punition, ce qui tend à montrer que l'influence de la sanction pourrait également s'inscrire dans une perspective dynamique.

Par ailleurs, l'influence de la récompense semble également présenter le défaut de s'effriter plus rapidement que celle de la punition, à l'approche de la fin du jeu. Selon nous, les comportements de « fin de jeu » sont dûs à un « relâchement » de certains des coopérateurs qui, lors des dernières périodes, diminuent leurs contributions car ils jugent avoir déjà consenti à un effort conséquent auparavant. Toutefois, d'autres coopérateurs, les plus assidus, ne font pas preuve de la même indulgence envers eux-mêmes et peuvent être prêts à punir ceux qui ne poursuivent pas l'effort jusqu'au bout. La présence de cette menace permet alors de limiter la survenance des effets de « fin de jeu » dans la condition « punition ». À l'inverse, dans la condition « récompense », un individu conscient que la fin du jeu approche et considérant avoir déjà fait suffisamment d'efforts, n'est probablement plus autant à la recherche de reconnaissance, de la part de ses partenaires, que lors des premières périodes et, de fait, se soucie moins de recevoir une récompense. De plus, le fait que les individus soient conscients qu'il ne seront bientôt plus en interaction les uns avec les autres, implique sans doute qu'il leur paraît moins intéressant de contribuer à construire de bonnes relations entre eux, ce qui limite l'usage de la récompense. Ainsi, ces deux phénomènes expliquent pourquoi, à l'approche de la fin du jeu, la capacité d'incitation du mécanisme de récompense diminue, provoquant ainsi une apparition plus précoces et prononcées des comportements de « fin de jeu ».

Face à ce problème, l'expérience de Rand, Dreber, Ellingsen, Fudenberg et Nowak (2009) permet, à travers un protocole original, de comparer l'effet de la punition et de la récompense, sur la coopération dans le jeu du bien public, sans que les perturbations dues aux comportements de « fin de jeu » n'interviennent. Pour ce faire, les participants à diverses variantes du jeu du bien public, se déroulant en « partner », sont informés que le jeu se joue de manière répétée, mais le nombre exact de répétitions ne leur est pas communiqué. Dans ces conditions, les auteurs montrent alors que les contributions au bien public déclinent dans la version standard du jeu, tandis que celles-ci se maintiennent à un niveau élevé lorsque l'option de punition ou de récompense est disponible pour les agents. Toutefois, les résultats suggèrent à nouveau que la moyenne des gains générés, à chaque période, en intégrant à la fois les gains liés à la phase de contribution et ceux issus de la phase d'interaction ciblée (punition ou récompense), sont plus importants dans la condition « récompense » que dans la condition « punition ». Par ailleurs, Rand *et al.* (2009) analysent également la fréquence d'utilisation de la récompense et de la punition, au cours du temps et leurs données montrent que, dans un tel contexte de jeu, l'utilisation de la punition tend à diminuer de période en période, tandis que celle de la récompense tend

à augmenter. Ce dernier résultat est cohérent avec la formation de liens sociaux positifs entre les agents de la condition « récompense ». En effet, ceux-ci se récompensent de plus en plus et tout se passe comme si une relation de qualité se créait entre eux, sur la base d'interactions réussies.

Cette étude laisse entrevoir les conditions permettant une utilisation optimale du mécanisme de récompense, à savoir, des individus qui interagissent de manière prolongée, sur une période indéterminée, dans un groupe restant inchangé et au sein duquel les individus sont identifiables (ce qui permet de cibler la récompense). Les résultats obtenus avec cette expérience sont d'autant plus pertinents que les conditions que nous venons d'énumérer se retrouvent dans de très nombreux environnements concrets, tels que le monde du travail, avec les contrats à durée indéterminée ou encore l'appartenance à une communauté de voisins, à une association etc. Dans chacun de ces contextes, les membres d'un même groupe interagissent ensemble de manière régulière, sont capables d'observer (dans une certaine mesure) les actions menées par leurs partenaires et font face à une incertitude quant à la durée de leur interaction ensemble. Chacune de ces situations se prête donc au recours du mécanisme incitatif de récompense afin de favoriser l'adoption de comportements coopératifs.

Au total, en ce qui concerne les possibilités de récompenser ou de punir, les résultats qui se dégagent expérimentalement sont en contradiction avec les prédictions du modèle classique. En effet, celles-ci sont largement utilisées par les joueurs lorsqu'elles sont disponibles et se traduisent par un impact significatif sur le niveau de contribution des agents, y compris lorsque les récompenses et les punitions sont immatérielles (Peeters et Vorsatz, 2013). Plus que l'aspect matériel, il semble que ce soit le fait d'offrir aux individus la possibilité d'une interaction ciblée (ne serait-ce que pour communiquer) avec leurs partenaires, qui soit un outil crucial dans la promotion et le maintien de la coopération (Rand *et al.*, 2009).

Par ailleurs, il est intéressant de noter qu'une méta-analyse largement documentée, menée par Balliet, Mulder et Van Lange (2011), a montré qu'il n'y a pas de différence significative d'efficacité entre récompense et punition, dans les situations de dilemme social, bien que la situation la plus favorable à la coopération soit celle où les partenaires disposent de ces deux possibilités (ce qui témoigne de leur complémentarité). Ce résultat, combiné avec le fait que la récompense permet d'obtenir des gains globalement plus importants que la punition et qu'il existe un lien entre émotions positives et bien-être (cf. 2.3.1), renforce l'intérêt quant à l'utilisation d'institutions permettant aux agents de se récompenser les uns les autres, que ce soit dans le jeu du bien public, ou, plus largement, dans les situations de dilemme social. A cet égard, l'étude de Sutter *et al.* (2010) démontre que lorsque les individus ont la possibilité de choisir le mécanisme incitatif disponible lors de leur participation au jeu du bien public, ceux-ci se tournent plus volontier vers l'usage de la récompense que vers le recours à un mécanisme punitif. Ainsi, bien que les résultats expérimentaux suggèrent que l'utilisation combinée de la punition et de la récompense soit la meilleure solution dans le but de promouvoir la coopération, à choisir, il semble que le planificateur social doive privilégier un mécanisme incitatif autorisant seulement la récompense (lorsque les conditions s'y prêtent) plutôt qu'une institution reposant sur la seule punition.

4.4. Régulation, communication et leadership

D'autres modifications de protocole, modifiant le déroulement du jeu du bien public, sont également apparues, au sein de la littérature, pour tenter de promouvoir la coopération.

Par exemple, l'expérience de Reeson et Tisdell (2008) examine le rôle de deux procédures modulant la participation à un jeu du bien public répété. D'une part, avant le début du jeu et entre deux périodes, plus tard dans le jeu, les sujets sont informés de l'aspect bénéfique de la contribution au bien public vis-à-vis du bien-être collectif, ce qui promeut la portée morale de la contribution. D'autre part, un mécanisme de régulation, consistant à contraindre le choix des participants, en imposant une contribution minimale, est également utilisé. Dans les faits, après avoir pu décider librement de leur contribution pendant un certain nombre de période, il est ensuite interdit aux sujets de réaliser une contribution inférieure à une valeur minimale (bien qu'ils puissent contribuer un montant supérieur à celle-ci) pendant une série de répétitions, avant que le choix de la contribution redevienne libre à nouveau. Les auteurs constatent alors que, dans la condition « information », où les sujets reçoivent l'information concernant les bienfaits de la coopération, les niveaux de contributions obtenus sont significativement plus importants que ceux observés dans la condition de contrôle (où les participants jouent à la version standard du jeu du bien public). En outre, par rapport à la condition « information », on observe, dans la condition où le mécanisme de régulation est implémenté, une chute drastique des niveaux de contributions, dès lors que la décision stratégique des agents redevient libre. On qualifie un tel phénomène d'effet d'éviction.

Ainsi, le recours à la régulation pose problème, en ce sens que celle-ci cloisonne (au moins en partie) le libre arbitre de l'individu et peut conduire à l'éviction de leur motivation intrinsèque à agir de manière prosociale. Les résultats de Reeson et Tisdell (2008) montrent ainsi qu'il est beaucoup plus efficace de convaincre les individus de l'intérêt de la coopération, en les laissant libre de leur choix, plutôt que de leur imposer la coopération, ne serait-ce que dans un but pédagogique. En effet, durant la procédure de régulation (pouvant être difficile et contraignante à mettre en œuvre dans les situations concrètes) les agents sont perturbés dans leur réflexion morale et peuvent perdre de vue leur motivation intrinsèque à contribuer, la seule raison justifiant leur coopération devenant alors le fait d'être contraint. Dès lors que la régulation cesse, la contrainte disparaît et, dans le cas où les sujets auraient abandonné une approche morale de la décision de contribution pour une approche matérielle, cela peut les conduire à être moins coopératifs par la suite. Cet effet pourrait également être en partie expliqué par le fait qu'avec l'arrivée de la régulation, les intentions des agents ne sont plus explicitement exprimées par leur décision de contribution, ce qui peut être un frein dans le développement des relations interpersonnelles, qui se fait habituellement sur la base des émotions suscitées par les actions entreprises dans le jeu. Lorsque la régulation est en place, les comportements deviennent en quelque sorte artificiels et ne sont probablement plus autant capables de déclencher l'émergence d'émotions positives, comme la gratitude, contribuant au renforcement des liens au sein des groupes, lesquels interviennent dans le maintien de la coopération. L'entretien de ces liens étant suspendu pendant que la régulation est en place, leur dégradation est envisageable, ce qui peut conduire à des comportements moins coopératifs à l'arrêt de la contrainte et plonger ensuite les groupes dans une dynamique de réciprocité négative, sous l'influence

d'émotions comme la colère ou la déception. Au contraire, offrir aux individus des outils permettant d'exprimer la survenance de leurs différentes émotions, à travers les actions de punitions et de récompenses ou, comme nous allons le voir, à travers la communication, permet de lubrifier les interactions sociales et de les rendre plus coopératives.

Comme nous l'avons vu, la transparence (identification des sujets et observabilité des comportements) est un élément crucial dans la promotion de la coopération, même lorsque les participants jouent anonymement. Une autre modification de protocole ayant été étudiée, s'agissant du jeu du bien public, est celle qui consiste à permettre aux participants de se rencontrer et/ou de communiquer (par écrit, à l'aide d'un enregistrement audio) ensemble, au cours du jeu.

Ainsi, dans l'expérience d'Isaac et Walker (1988), des joueurs ayant participé à plusieurs répétitions du jeu du bien public, se rencontrent en face à face quelques minutes, avant chaque nouvelle répétition du jeu, afin de pouvoir discuter des niveaux de contribution qu'il convient de mettre en œuvre. Cette discussion n'engage en rien les individus qui restent libres au moment de leur choix et cela ne modifie pas la structure des gains dans le jeu, ce qui implique que l'équilibre de Nash reste inchangé et que chaque joueur a toujours intérêt à ne pas contribuer. Malgré tout, un tel procédé permet une augmentation significative des niveaux de contribution. De plus, lorsque la séquentialité du déroulement de l'expérience est inversée, avec une première série de répétitions où les agents peuvent communiquer et une seconde durant laquelle cette possibilité est enlevée, les niveaux de contributions restent très élevés.

Par ailleurs, l'expérience de Brosig, Weimann et Ockenfels (2003) examine un certain nombre de modes de communication et d'identification des sujets et les données obtenues indiquent que le mode de communication le plus efficace est celui de la rencontre en face à face, laquelle permet une communication verbale et visuelle, en plus d'offrir une identification explicite des sujets. L'analyse du contenu des échanges a également permis de mettre en évidence le fait que les individus recherchent collectivement la marche à suivre pour maximiser leurs gains, ce qui les conduit généralement à la détermination de l'optimum du jeu (où tous les joueurs contribuent au maximum de leur dotation). Les participants cherchent alors souvent à s'engager les uns auprès des autres pour tendre vers une coopération maximale. Certains sujets ont également mis en exergue le fait que tout manquement à cet engagement serait sanctionné par des représailles consistant en un arrêt de la coopération de leur part. Bien qu'une nouvelle fois, ces engagements n'aient pas de nature contractuelle assurant leur respect, les résultats montrent que les sujets ont, la plupart du temps, tenu leurs promesses, y compris en ce qui concerne la mise en œuvre des représailles. A cet égard, les résultats de cette étude sont cohérents avec ceux obtenus dans l'expérience de Masclet, Noussair et Villeval (2013), au cours de laquelle les sujets doivent communiquer le montant de punition qu'ils ont l'intention d'imposer, pour chaque niveau de contribution possible. Dans les faits, ces annonces ont globalement permis une coopération supérieure, par rapport au jeu joué sans communication et des punitions ont été effectivement imposées par certains sujets.

Selon nous, l'intérêt de la communication entre les agents tient à deux éléments. Premièrement, c'est un formidable outil de coordination, qui permet aux agents de se mettre d'accord sur un comportement à adopter, ce qui les conduit souvent à prendre un engage-

4. *Le jeu du bien public : un cadre approprié à la représentation du rôle économique de la gratitude*

ment les uns envers les autres. Comme nous avons déjà pu en discuter, promesses faites par les individus ont une influence reconnue sur leur comportement et cela en fait un moyen relativement efficace dans le but de promouvoir la coopération. Par ailleurs, le fait même de laisser aux agents l'opportunité d'échanger leurs points de vue sur la bonne marche à suivre, leur permet de savoir à quoi s'attendre, ainsi que de mieux anticiper les conséquences de leurs actes en ce qui concerne les jugements et les émotions qu'ils susciteront chez autrui. Ainsi, la communication est susceptible de limiter l'émergence d'émotions à valence négative, comme la colère ou la déception, puisque les agents connaissent en détail les attentes des autres et sont conscients du fait que leurs comportements révèlent plus clairement leurs intentions face à ces mêmes attentes.

Deuxièmement, nous pensons que la communication est importante car elle est susceptible de promouvoir la création et le renforcement des liens sociaux. Lorsque des individus se rencontrent, le fait qu'ils puissent échanger et discuter ensemble, le fait qu'ils apprennent à se connaître, crée naturellement un lien entre eux qui est susceptible de les orienter vers une plus grande coopération.

Toutefois, selon Villeval (2010), l'efficacité de la communication peut être sensible à des facteurs comme un accroissement de la taille du groupe ou l'introduction d'asymétries au niveau des dotations. Dans ces différents cas, la communication peut ralentir la décroissance du niveau de contribution, mais ne parvient plus à l'enrayer.

Enfin, la séquentialité est un autre type de modulation du déroulement du jeu du bien public ayant été envisagé comme un moyen de favoriser la coopération. En effet, lorsque le jeu standard est modifié de sorte à ce qu'un joueur puisse jouer avant les autres membres de son groupe, en étant observé par ses partenaires, celui-ci a la capacité de pouvoir influencer leurs décisions. Cette modification permet de représenter l'existence dans les groupes humains de « leaders », qui n'hésitent pas à être à l'initiative et peuvent emporter les partenaires avec eux dans leur sillage. Ainsi, lors d'une campagne de don, le fait de citer en exemple, auprès de bienfaiteurs potentiels, un donateur ayant fortement contribué, permet d'augmenter significativement leurs dons (Croson et Shang, 2008). Ce phénomène peut être expliqué par l'existence de préférences pour le conformisme chez certains agents, par le fait que les êtres humains ont une tendance naturelle à l'imitation (qui est un mode d'apprentissage très important) et également parce qu'un comportement coopératif, de la part du premier joueur, permet d'envoyer aux coopérateurs conditionnels le signal de la présence d'un individu prêt à coopérer, ce qui peut rassurer ces derniers dans leur décision (Villeval, 2010).

L'influence du leadership et des processus d'imitation est d'autant plus important que les comportements de coopération, comme les comportements non coopératifs, sont capables de se transmettre au sein des réseaux d'agents, et de persister durant plusieurs périodes. Par exemple, Fowler et Christakis (2010) montrent, dans le cadre d'un jeu du bien public répété, joué en « stranger », que l'action de certains individus est capable d'avoir une influence sur le comportement des joueurs qui se retrouvent dans le même groupe, lors des périodes suivantes (alors même qu'ils ne jouent plus ensemble). Les résultats de Fowler et Christakis (2010) suggèrent ainsi que chaque contribution additionnelle au bien public, réalisée à la première période, est triplée, au cours de l'expérience, par l'augmentation de la contribution d'autres sujets, ayant été directement ou indirectement

influencés. Ce résultat témoigne de l'extrême importance des comportements réalisés au départ du jeu et de la nécessité d'avoir des éléments moteurs dans les groupes, de sorte à permettre une stimulation dynamique la coopération, à travers leur exemple. Toutefois, cette méthode peut être compliquée à mettre en œuvre, en pratique, puisqu'elle repose sur le comportement du « leader » et que cela induit qu'il supporte un coût individuel relativement élevé.

Au total, de nombreux moyens existent pour lutter contre les comportements égoïstes et soutenir la coopération dans les situations de dilemme social de type jeu du bien public. Généralement, il convient de combiner différents mécanismes pour obtenir le meilleur résultat possible. Ainsi, pour que la coopération ait le plus de chance d'être initiée et maintenue (voire promue), il est préférable que les agents soient identifiables, que leurs comportements soient parfaitement observables par l'ensemble des membres de leur groupe, que les individus puissent se rencontrer et discuter afin de se coordonner sur un comportement de référence, qu'un individu puisse éventuellement donner le ton en jouant avant les autres, que les groupes restent inchangés au cours du temps (condition « partner ») et que les individus aient l'opportunité d'une interaction ciblée avec chacun des membres de leur groupe (communication verbale, punition, récompense), en réponse aux décisions de contribution.

4.5. Les atouts du jeu du bien public

Nous avons fait le choix de placer notre analyse dans le contexte particulier du jeu du bien public pour de multiples raisons.

Tout d'abord, le jeu du bien public est représentatif d'une multitude de situations économiques, touchant à des problématiques comme la protection de l'environnement, l'organisation des entreprises, l'évasion fiscale, l'abstentionnisme etc. Des exemples concrets de situations pouvant être représentées par un tel jeu sont, entre autres, le fait de payer pour les transports en commun, de contribuer à un projet de groupe ou de se livrer à un travail d'équipe (Balliet, Mulder et Van Lange, 2011 ; Gächter et Herrmann, 2009). Chacun d'entre nous est confronté quotidiennement à de telles situations et la coopération collective des agents y est souhaitable de façon à promouvoir le bien-être social. Étudier les ressorts de la coopération dans le jeu du bien public est donc un enjeu de société primordial.

De plus, dans la perspective de la modélisation du rôle économique des émotions morales, et notamment de la gratitude, la richesse du jeu du bien public réside dans le fait que l'ensemble des joueurs a la possibilité d'être bienveillants (malveillants) avec les autres membres de son groupe, ainsi que de faire preuve de réciprocité positive (négative) à leur égard. En effet, cela est possible, d'une part, à travers la décision de contribution au bien public, consentie par chaque joueur, et, d'autre part, cela est également permis par l'usage de l'option de récompense (punition) coûteuse, lorsque celle-ci est à la disposition des agents. Ainsi, la structure du jeu est favorable à l'analyse des principales tendances à l'action des émotions d'évaluation d'autrui que sont la colère, la déception ou, plus particulièrement, la gratitude et, contrairement à une situation comme celle du jeu du dictateur, par exemple, le fait que chaque agent soit amené à prendre une décision stratégique offre une grande richesse d'interaction entre les individus.

4. Le jeu du bien public : un cadre approprié à la représentation du rôle économique de la gratitude

En outre, le jeu du bien public peut être joué par plus de deux joueurs. Cet élément est particulièrement pertinent, compte tenu du fait que l'influence d'émotions comme la gratitude, sur les relations interpersonnelles, peut être perçue à la fois au niveau dyadique, comme à l'échelle du groupe et que celle-ci peut se propager au sein d'un réseau d'agents.

Enfin, le jeu du bien public peut être joué de manière répétée, ce qui est opportun vis-à-vis de l'aspect dynamique du fonctionnement de l'influence des émotions morales sur la qualité des relations entre les agents et le développement du lien qui les unit au cours d'une interaction prolongée.

Pour toutes ces raisons, le jeu du bien public semble être le cadre idéal pour modéliser l'influence des émotions morales sur le comportement des individus, dans le contexte des dilemmes sociaux. Par ailleurs, le fait que ce jeu ait été étudié dans de nombreuses expériences, en économie expérimentale et en psychologie sociale, nous garantit l'accès à des résultats expérimentaux abondants nous offrant la possibilité d'examiner la validité et la pertinence des résultats de notre modèle.

Quatrième partie .

Le Rôle des Emotions Morales dans l'Analyse Formelle du Jeu du Bien Public avec Phase de Récompense

5. Le Cas Statique

5.1. Préambule au modèle : Structure du jeu du bien public

Par souci de clarté et de concision, nous procédons ici, en préambule à la présentation de notre modèle, à une exposition détaillée de la structure du jeu du bien public, qui sert de cadre d'analyse à la formalisation. En effet, nous étudions une multitude de situations, à partir de différentes déclinaisons du jeu du bien public ayant une base commune. Il est donc préférable de décrire ici la structure du jeu, de manière complète et relativement générale, afin que le lecteur puisse avoir ensuite plus de facilité pour appréhender la pluralité des situations théoriques que nous envisageons.

La situation la plus générale que nous examinons est celle du jeu du bien public avec phase de récompense, joué par un groupe de $n > 2$ agents. Par conséquent, nous présentons ici cette situation qui englobe toutes les déclinaisons possibles du jeu utiles à l'analyse.

5.1.1. Structure formelle du jeu

5.1.1.1. Présentation générale

Préalablement au commencement du jeu, on suppose que chaque agent est assigné, avec ses partenaires, à un groupe composé de n joueurs et que chaque partenaire se voit attribuer une dotation initiale identique. Le jeu du bien public (JBP), avec phase de récompense, se déroule ensuite en deux étapes. La première phase correspond à une phase de contribution durant laquelle, chaque joueur doit décider, simultanément à ses partenaires, du montant qu'il souhaite utiliser, à partir de sa dotation initiale, pour contribuer au financement d'un bien public qui bénéficie ensuite à l'ensemble des agents du groupe.

A la fin de la première étape, chaque joueur récupère les gains issus de la phase de contribution et est informé de la décision stratégique de ses partenaires. Sur la base de cette information, chaque joueur doit ensuite choisir, au cours de la deuxième étape du jeu, un montant de récompense, qui peut être positif ou nul, à destination de chacun de ses partenaires. Par ailleurs, on suppose que l'acte de récompenser est coûteux et, plus précisément, que chaque agent supporte un coût proportionnel au montant de sa récompense.

5. Le Cas Statique

5.1.1.2. Déroulement du jeu et notations

On note $I = \{1, 2, \dots, n\}$ l'ensemble des joueurs.

5.1.1.2.1. La phase de contribution

Pour jouer, chaque agent $i \in I$ est pourvu d'une dotation initiale notée $d_i = d$.

Lors de la première phase, chaque joueur $i \in I$ doit donc décider de sa contribution au bien public $y_i \in [0, d]$. Lorsque l'ensemble des joueurs a contribué, chaque membre du groupe reçoit la somme $a \times (\sum_{i \in I} y_i)$, provenant du montant total investi dans le bien public, où a correspond au rendement marginal du bien public. Par ailleurs, on suppose que $\frac{1}{n} \leq a \leq 1$, afin de s'assurer que le jeu ait bien les propriétés d'un dilemme social.

Pour tout $i \in I$, le gain matériel de l'agent i est défini comme :

$$\pi_i = d - y_i + a \left(\sum_{i \in N} y_i \right) \quad (5.1.1)$$

La phase de contribution s'achève par l'obtention, pour chaque joueur, de son gain monétaire.

Le jeu du bien public, dans sa version *standard*, se limite à cette seule phase de contribution. La dénomination *standard* renvoie à l'idée que cette version du jeu est la première qui soit apparue historiquement, comme un dérivé du jeu du dilemme du prisonnier.

Lorsque le jeu se limite à la seule phase de contribution, étant donné que $\frac{1}{n} \leq a \leq 1$, la stratégie dominante, pour un joueur rationnel et égoïste, est de ne rien contribuer, même si, collectivement, l'optimum social est atteint lorsque chaque joueur contribue l'intégralité de sa dotation.

Nous présentons maintenant la deuxième phase de notre jeu, qui offre à chaque agent la possibilité de récompenser chacun de ses partenaires.

5.1.1.2.2. La phase de récompense

A la fin de la première étape, chaque joueur est informé de la décision stratégique de chacun de ses partenaires. Sur la base de cette information, chaque agent $i \in I$ doit ensuite choisir, au cours de la deuxième étape du jeu, un montant r_{ij} de récompense, qui peut être nul, à destination de chacun des autres joueurs $j \in I, j \neq i$.

Par ailleurs, lorsqu'un joueur récompense un partenaire, celui-ci supporte un coût unitaire noté c . De plus, chaque agent i peut potentiellement recevoir une récompense r_{ji} de la part de tout autre agent $j \in I, j \neq i$.

Finalement, pour tout $i \in I$, $i \neq j$, le gain monétaire obtenu à la fin du jeu du bien public avec phase de récompense est égal à :

$$\Pi_i = \pi_i - c \sum_{j \neq i} r_{ij} + \sum_{j \neq i} r_{ji} \quad (5.1.2)$$

Sachant que π_i correspond aux gains liés à la phase de contribution (décrits dans l'équation (5.1.1)).

Puisqu'il est coûteux d'offrir une récompense à un partenaire, la meilleure stratégie, pour un joueur égoïste et rationnel, est de s'abstenir de tout niveau positif de récompense. En procédant à un raisonnement par induction à rebours, il vient que l'introduction d'une seconde phase, offrant une opportunité de récompense, n'est pas en mesure de changer le comportement d'agents individualistes et calculateurs lors de la phase de contribution (puisque cette opportunité n'est pas utilisée en deuxième étape), et finalement, le modèle classique, en théorie des jeux, suppose que personne ne contribue (ni ne récompense) à l'équilibre de Nash.

Comme nous avons pu le voir lors de ce préambule, une représentation des agents économiques comme des individus rationnels et qui valorisent uniquement leurs propres gains matériels conduit à un résultat théorique en contradiction avec les données empiriques qui ont été dégagées concernant le jeu du bien public, s'agissant de sa version standard ou lorsque celui-ci est joué avec une phase subséquente de récompense entre les partenaires (cf. 4.2).

Dans la suite de ce chapitre, nous allons voir comment notre représentation originale du processus de décision des agents est capable de produire des résultats qui sont davantage en adéquation avec ce qui est observé empiriquement, via l'introduction du rôle des sentiments moraux.

5.2. Le Modèle basique à deux joueurs

Comme nous l'avons vu précédemment, l'analyse formelle des comportements économiques, dans le contexte du jeu du bien public, a déjà été effectuée à la lumière du rôle influent des émotions. En particulier, le modèle de Bowles et Gintis (2005) (cf. 3.3.2.1) est centré sur l'intégration des deux émotions morales de culpabilité et de honte pour décrire le processus de décision des agents lorsqu'ils jouent à un jeu du bien public, incluant une première phase de contribution, suivie d'une phase de punition.

Bien qu'étant les seules émotions explicitement prises en compte dans la forme fonctionnelle développée par les auteurs, celle-ci englobe également des préférences en lien avec des considérations de réciprocité, ce qui fait écho au rôle d'autres émotions comme la juste colère et la gratitude. En outre, ce modèle s'articule également autour de l'existence de normes sociales et morales dans l'économie, en analysant la manière dont les agents réagissent face à la présence de telles normes de comportement.

Selon nous, s'agissant du contexte des dilemmes sociaux, ce modèle pose solidement les bases de l'intégration de l'influence des émotions morales et prosociales, dans la représentation du processus de décision des agents économiques. Le travail de modélisation présenté dans le cadre de cette thèse tente de s'inscrire dans la lignée de la contribution de Bowles et Gintis (2005), en ouvrant l'analyse à des situations qui mettent en jeu l'émergence d'émotions à valence positive, comme la gratitude.

En guise d'introduction, nous commençons par l'analyse de la version *standard* du jeu du bien public (JBP), joué en un coup, par *deux* agents. Pour rappel, la dénomination *standard* est utilisée pour désigner le JBP lorsqu'il se compose de la seule phase de contribution. Avant de passer au cas du JBP avec phase de récompense ultérieure, cette première étape nous permet d'entamer la présentation formelle des composants de notre fonction d'utilité, en commençant par la situation la plus simple possible. En outre, celle-ci nous offre aussi l'occasion d'examiner, par comparaison, la manière dont notre modèle est en mesure d'expliquer comment l'ajout d'une phase de récompense influence les comportements individuels vers la promotion de la coopération dans le JBP.

5.2.1. Le jeu du bien public dans sa version standard

Nous considérons d'abord le jeu du bien public *standard*, joué en un coup, par $n = 2$ joueurs.

L'ensemble des joueurs se limite donc ici à $I = \{i, j\}$, avec $i \neq j$.

Nous rappelons que chaque joueur $i \in I$ a une dotation de $d_i = d$ unités monétaires et qu'il doit simultanément décider du montant $y_i \in [0, d]$ qu'il souhaite contribuer au bien public, sachant que chaque joueur en retire $a \sum_i y_i$ unités monétaires, avec $\frac{1}{2} \leq a \leq 1$.

Le gain monétaire de chaque joueur $i \in I$ est alors donné par :

$$\pi_i = d - y_i + a(y_i + y_j) \tag{5.2.1}$$

La version *standard* du JBP n'est pas examinée dans l'article de Bowles et Gintis (2005). Ainsi, bien que notre formalisation se calque sur la fonction d'utilité des auteurs et que nous soyons en accord avec plusieurs de leurs hypothèses, concernant la représentation du processus de décision des individus, notre travail permet néanmoins d'apporter de la nouveauté en explorant de nouvelles interprétations des différentes composantes de la fonction, ainsi qu'en approfondissant la compréhension des mécanismes psychologiques qui les sous-tendent, tout en fournissant une prédiction théorique du comportement des agents dans cette version du jeu.

Pour l'instant, dans ce cadre de jeu très simple, notre fonction d'utilité saisit quatre objectifs principaux affectant le processus de décision des individus : (1) leurs propres gains matériels dans le jeu et (2) l'intensité avec laquelle ils se sentent coupables, du fait de leurs propres actions ; ainsi que la manière dont ils valorisent les gains de leurs partenaires, qui dépend du (3) capital de sympathie (que nous expliquons ci-après) dont les joueurs

disposent à l'égard de leur partenaire, ainsi que de (4) leurs ressentis émotionnels suite à l'évaluation du comportement de ce dernier.

5.2.1.1. La fonction d'utilité

Pour tout agent $i \in I$, $i \neq j$, la fonction d'utilité est définie comme :

$$U_i = AC_i + EA_{ij} \quad (5.2.2)$$

Les notations AC et EA se réfèrent à la typologie des émotions morales qui est détaillée dans le premier chapitre de ce travail. En effet, la notation AC se rapporte aux émotions dites auto-conscientes, pour lesquelles le focus de l'agent est tourné vers lui-même, tandis que la notation EA est utilisée en référence aux émotions d'évaluation d'autrui, qui interviennent lors d'une interaction entre l'individu et un autre agent.

Le terme AC_i est défini, pour tout agent $i \in I$, de la manière suivante :

$$AC_i = \pi_i - \beta_i(\bar{y}_i - y_i)^2$$

Où $\beta_i \geq 0$ et $\bar{y}_i \in [0, d]$

Pour tout agent $i \in I$, $i \neq j$, le terme EA_{ij} est égal à :

$$EA_{ij} = \theta_{ij}\pi_j$$

Avec θ_{ij} , le poids accordé aux gains de j , égal à :

$$\theta_{ij} = \delta_i + \alpha_i(y_j - \bar{y}_i) \quad (5.2.3)$$

Où $\alpha_i \geq 0$.

Au total, on peut réécrire la fonction d'utilité de l'agent i sous la forme suivante :

$$U_i = \pi_i - \beta_i(\bar{y}_i - y_i)^2 + \theta_{ij}\pi_j \quad (5.2.4)$$

Enfin, il est à noter que si l'on pose $\alpha_i = \delta_i = \beta_i = 0$, nous retrouvons la représentation classique des agents en théorie des jeux, qui peut alors être considérée comme un cas particulier de notre formalisation.

5.2.1.1.1. Emotion auto-consciente et le terme de culpabilité

Tout d'abord, de manière similaire au modèle de Bowles et Gintis (2005), nous supposons que chaque agent $i \in I$ est averse au fait d'adopter un comportement qui viole la norme de contribution \bar{y}_i que celui-ci a intériorisée.

Cette norme \bar{y}_i , peut être envisagée comme étant liée à ce que Yamagishi et Mifune (2008) décrivent comme une « heuristique de groupe », *i.e.* une *décision stratégique adoptée par défaut* par les individus, lorsqu'ils font face à une situation de groupe.

D'après Yamagishi et Mifune (2008), ce qui pousse originellement les agents à recourir à une stratégie par défaut, face à une situation de groupe, est le fait que celle-ci est susceptible de faciliter leur admission au sein d'un groupe. En effet, en supposant que ce groupe fonctionne sur la base d'un système d'échange généralisé (où la réciprocité indirecte a une place importante), il est alors stratégique pour les individus de coopérer, de façon à pouvoir ensuite bénéficier (et à ne pas être privés) des bienfaits réservés à ceux qui appartiennent au groupe. En particulier, être bienveillant à l'égard des autres membres d'un tel groupe permet d'éviter d'être catégorisé comme un « passager clandestin » avant d'être ensuite exclu du système. « L'heuristique de groupe » suppose alors que les agents font l'hypothèse, par défaut ¹, que lorsqu'ils font face à une situation de groupe, celui-ci fonctionne sous la forme d'un système d'échange généralisé. En conséquence, les individus ont une incitation à agir de manière altruiste et coopérative, de sorte à limiter le risque de développer une mauvaise réputation auprès des autres membres du groupe.

Dans cette perspective, nous supposons que les individus développent de telles « heuristiques de groupe », sur la base de l'éducation morale qu'ils reçoivent, ainsi que de leurs précédentes expériences d'interaction sociale. De plus, nous supposons également que celles-ci fonctionnent ensuite comme des points de référence pour guider et juger leurs comportements (*i.e.* des normes), dont ils se servent, par réflexe, dès qu'ils doivent interagir avec autrui (peu importe le nombre d'agents impliqués). Enfin, on considère que, quel que soit le contexte auquel les agents font face, ils sont capables de déterminer « l'heuristique de groupe » qui s'adapte à la situation en question et qui est en accord avec leurs principes moraux. Ainsi, on suppose que, dans le contexte du jeu du bien public, chaque agent $i \in I$ est capable de déterminer la norme de contribution \bar{y}_i qu'il convient moralement de jouer selon lui.

Formellement, en raison de la présence du carré, $\beta_i(\bar{y}_i - y_i)^2$ exprime toujours un coût psychologique, supporté par les agents lorsqu'ils choisissent un niveau de contribution qui diffère de leur norme, où β_i est un paramètre qui en mesure l'intensité.

Ainsi, lorsque $y_i < \bar{y}_i$, notre modèle suppose que le joueur i ressent de la culpabilité du fait de ne pas parvenir à contribuer à hauteur de ce qu'il considère comme le comportement socialement ou moralement approprié. Réaliser une contribution inférieure à \bar{y}_i constitue une violation de la norme intériorisée par i qui, de surcroît, représente un coût pour autrui. En effet, une telle violation entraîne des retombées, provenant du bien public, qui sont moindres, pour l'ensemble des joueurs. En ce sens, une telle action est typiquement susceptible d'être à la source de l'émergence de la culpabilité (cf. 1.2.2.2.1).

1. Le fait de supposer que les individus font l'hypothèse par défaut d'un système d'échange généralisé permet d'expliquer les niveaux de coopération observés dans les jeux de dilemme social joués en un coup.

Toutefois, le recours à cette spécification particulière implique que les agents se sentent coupables, même dans le cas où leur contribution dépasse la valeur de leur norme, ce qui peut sembler contre-intuitif. Parallèlement à la justification judicieuse fournie par Bowles et Gintis (2005), que l'on a examinée plus tôt (cf. 3.3.2.1), nous fournissons également ici une approche complémentaire pour expliquer ce phénomène.

En particulier, il semble pertinent de relier le coût psychologique découlant du fait de contribuer plus que ce qui est recommandé par la norme, à la crainte d'être exploité par les autres en agissant de la sorte. En effet, Vohs, Baumeister et Chin (2007) définissent le sentiment d'avoir été dupé comme une réponse émotionnelle aversive à la perception que l'on a été exploité, durant une transaction interpersonnelle (principalement lorsque cela implique un échange économique), en partie en raison de ses propres décisions. Décrire les agents comme relativement indisposés par le fait de contribuer trop, par rapport à leur propre norme de comportement, semble être une manière appropriée de rendre compte de cet état aversif. En parallèle, les agents pourraient également souffrir de la peur d'être considérés comme des imbéciles (ou des idiots) en se mettant délibérément dans une situation désavantageuse. Auquel cas, la culpabilité ressentie pourrait être interprétée comme liée à un heurt qu'ils s'imposent à eux-mêmes, à travers un acte délibéré, les conduisant à se démunir. Cette conception est cohérente avec le fait que Burton-Chellew et West (2013) aient montré expérimentalement qu'une présentation du jeu du bien public mettant l'accent sur la manière dont les partenaires d'un individu bénéficient de sa contribution implique une diminution de son niveau de coopération. En effet, ce cadrage spécifique peut être perçu, par les agents, comme un étalage de l'ampleur avec laquelle les autres sont susceptibles de profiter (d'abuser) d'eux, lorsqu'ils contribuent. Notre modèle est donc en mesure de représenter un tel phénomène, à travers une valeur plus importante de β_i , dans le cas où l'agent i serait inséré dans ce contexte particulier.

En outre, il est également possible qu'un niveau démesuré de contribution puisse être interprété comme une volonté de se mettre en valeur et d'attirer l'attention d'autrui. Dans la plupart des interactions sociales, la vantardise est mal considérée au sein des groupes d'individus et peut entraîner une certaine désapprobation. Un tel phénomène correspond par exemple au postulat de Bénabou et Tirole (2006) qui affirment que le fait de vouloir se montrer peut heurter l'estime de soi des autres ou renvoyer l'image de quelqu'un d'irresponsable. Par conséquent, si l'intériorisation d'une norme est comprise comme un moyen, pour les agents, de s'intégrer dans les groupes sociaux, à travers la mise en œuvre de comportements respectueux envers les autres, il devient plus aisé de comprendre pourquoi la norme de contribution, entre autres, doit être respectée à la lettre. Ainsi, dans le JBP, le respect de la norme est nécessaire, d'une part, pour empêcher l'expression d'une désapprobation sociale consécutive à l'adoption d'un comportement égoïste et, d'autre part, comme un moyen d'éviter de subir l'expression d'un dénigrement social, résultant du fait de s'être mis en avant. C'est pourquoi, on peut considérer que l'émotion de culpabilité a évolué de sorte à guider les agents en les aidant à rester à l'écart de ces deux types de comportements socialement inappropriés.

Au delà du fait d'être plutôt convaincante, cette dernière interprétation a aussi l'avantage d'être compatible avec certains résultats empiriques relatifs au JBP avec phase de punition, qui montrent que, de temps à autres, des agents ayant contribué davantage que

5. Le Cas Statique

les autres, au bien public, sont la cible de ce qui peut être qualifié de *punitions malveillantes* de la part de leurs partenaires dans le jeu (Fehr et Gächter, 2000).

Néanmoins, cette interprétation exige que lorsqu'un joueur contribue trop, son partenaire puisse en ressentir de la désutilité, ce qui est impossible dans notre modèle. En conséquence, la forme fonctionnelle que nous utilisons ici ne se conforme pas pleinement à une telle explication.

Au total, la manière dont nous concevons, dans notre modèle, l'émergence et le rôle de l'émotion de culpabilité, ainsi que l'importance de la norme de contribution dans la décision stratégique des agents, se rapproche du fonctionnement de « l'heuristique de groupe ». On peut ainsi envisager la norme de contribution internalisée par les agents comme leur propre point de vue (développé à travers leur éducation et leurs expériences) à propos du comportement « type » à adopter, par défaut, de manière à être accepté par autrui lors de la participation à un jeu du bien public. L'émotion de culpabilité et en particulier l'anticipation de son émergence, leur permet alors, au moment de la prise de décision, d'être guidés vers ce point de référence. En ce sens, il semble que l'intérêt de l'émotion de culpabilité, en termes de valeur sélective, soit de guider les agents, à travers son rôle dans l'activation des « heuristiques de groupe », vers l'adoption de comportements moralement appropriés leur permettant d'être mieux intégrés socialement et acceptés par autrui.

5.2.1.1.2. La prise en compte d'autrui

La composante clé de la fonction d'utilité est θ_{ij} (équation 5.2.3). Ce terme nous permet d'introduire, dans le processus de décision des agents, la relation à autrui. Plus précisément, il représente la manière dont les individus valorisent le fait de percevoir leur partenaire dans le jeu comme un prolongement d'eux-mêmes, au sens où ils prennent en considération le bien-être matériel de ce dernier. Ainsi, une telle spécification pose implicitement que, pour chaque agent, la présence de l'autre, du fait de l'interaction qui a lieu avec lui, est appréhendée à travers les gains matériels obtenus par ce dernier. En d'autres mots, étant donné que les gains matériels d'un joueur constituent, en partie, son bien-être et qu'ils sont clairement identifiables dans le jeu, ils sont utilisés comme un indice de la satisfaction d'un partenaire, au moment de la prise de décision. Nous poursuivons avec la présentation formelle du terme θ_{ij} .

Le capital de sympathie

Pour commencer, nous nous concentrons sur δ_i , un paramètre pouvant paraître anodin de prime abord et dont nous considérons que le rôle a pu être sous-estimé dans l'article de Bowles et Gintis (2005). Ici, plutôt que de considérer que, pour tout agent $i \in I$, δ_i représente une force motrice altruiste et inconditionnelle vis-à-vis de l'autre agent, nous concevons plutôt δ_i comme le capital de sympathie que l'agent i accorde à son partenaire. De manière symétrique, on peut également envisager δ_i comme le capital de sympathie dont j dispose auprès de i .

La sympathie peut se définir comme un penchant naturel, spontané et chaleureux de quelqu'un envers une autre personne. Cette définition illustre bien notre conception du paramètre δ_i .

Enfin, nous verrons dans la suite de l'analyse (cf. chapitre 6) que ce paramètre individuel est amené à évoluer dans le temps et nous considérons ainsi qu'il correspond à un indice représentatif de la *perception* de l'agent i à l'égard du niveau de qualité de la relation qu'il entretient avec son partenaire, dans le cadre de cette interaction. Cet indice englobe notamment le niveau de proximité ou d'intimité que l'individu i ressent vis-à-vis de son partenaire. Au demeurant, l'intimité a été dégagée comme une des dimensions principales de la qualité d'une relation (Hassebrauck et Fehr, 2002).

Cette réinterprétation originale du paramètre δ_i , offre des perspectives intéressantes quant à l'interprétation du fonctionnement de ce modèle. En effet, il convient d'examiner ici la signification du paramètre δ_i , en fonction de son signe.

Lorsque $\delta_i > 0$, nous supposons qu'une valeur plus importante de δ_i implique que l'agent i se sent plus proche (intime) de son partenaire j ($i \neq j$), qu'il développe une perception plus positive du fait d'interagir avec lui et, finalement, il exhibe une tendance inconditionnelle plus importante à éprouver de la sympathie pour j , ce qui le motive à valoriser d'autant plus positivement le sort de son partenaire dans le jeu.

Parallèlement, lorsque $\delta_i = 0$, l'agent i ne se sent ni réellement proche de son partenaire, ni foncièrement distant (socialement) vis-à-vis de lui. Il a alors une perception relativement neutre du fait d'interagir avec son partenaire, et, en fin de compte, n'exhibe pas de tendance inconditionnelle particulière à éprouver de la sympathie pour j , ce qui n'entraîne aucune influence sur la manière dont i valorise le bien-être matériel de son partenaire dans le jeu.

Enfin, lorsque $\delta_i < 0$, une valeur plus importante de δ_i , en valeur absolue, implique que l'agent ressent un éloignement (social) d'autant plus important vis-à-vis de l'autre joueur, qu'il développe une perception d'autant plus négative du fait d'avoir à interagir avec lui et, au total, qu'il exhibe une tendance inconditionnelle d'autant plus grande, à éprouver de l'antipathie (ou aversion) à l'encontre de son partenaire, ce qui le pousse à valoriser négativement le sort de ce dernier.

Pour résumer, nous définissons δ_i comme un indice de la sympathie (antipathie) que i ressent vis-à-vis de j , qui est représentatif de l'intensité avec laquelle i se sent proche (distant) socialement de l'agent j , ainsi que du niveau de confiance qu'il accorde à ce dernier. De plus, on suppose que ce niveau de proximité et de confiance conduit i à développer une tendance *inconditionnelle*, d'intensité comparable, à valoriser positivement (négativement) le sort matériel de j dans le jeu.

Néanmoins, on peut s'interroger sur la façon dont se détermine le capital de sympathie qu'un individu accorde à un autre, lorsque les deux agents se rencontrent pour la première fois, autrement dit avant la survenance d'un quelconque historique d'interactions entre eux. Par souci de simplicité, nous supposons ici que, pour chaque agent i , le capital de sympathie accordé à toute personne anonyme, avec qui i doit interagir, est le même, quelle que soit la personnalité spécifique (en termes de paramètres) qui caractérise l'individu auquel l'agent i fait face².

2. Cependant, une façon plus réaliste, mais sans doute trop sophistiquée dans le cadre de ce modèle, de représenter ce capital, pourrait être de le faire dépendre (éventuellement par des processus d'analogie)

5. Le Cas Statique

Une telle conceptualisation de ce paramètre se rapproche de la notion « *d'orientation sociale* », que l'on trouve originellement dans des travaux de psychologie sociale, avant d'être ensuite adoptée par certains économistes. Ainsi, « *l'orientation sociale* » d'un individu correspond à l'attitude que celui-ci adopte, globalement, face à une personne anonyme (van Dijk, Sonnemans et van Winden, 2002 ; Sonnemans, van Dijk et van Winden, 2006).

Expérimentalement, pour mesurer cette attitude, les psychologues utilisent la méthode du test de l'anneau (Liebrand, 1984) qui conduit à classer les individus au sein de cinq catégories. Ces différentes catégories, ainsi que leur définition, sont indiquées dans le tableau suivant :

Catégorie d'orientation sociale	Définition
Agresseur	Qui souhaite le pire pour l'autre, quel que soit le résultat pour lui-même
Compétiteur	Qui veut être mieux loti que son vis-à-vis
Individualiste	Qui essaye d'obtenir le meilleur résultat possible pour lui-même
Coopérateur	Qui vise le meilleur, à la fois pour lui-même et pour l'autre
Altruiste	Qui désire le meilleur pour l'autre, quel que soit le résultat pour lui-même

TABLE 5.1.: Les différentes catégories « d'orientation sociale »

Ainsi, on peut supposer que la valeur δ_i qui caractérise un agent dans le modèle est illustrative de son « orientation sociale ». Le tableau suivant fournit des informations illustratives sur la valeur prise par le paramètre δ , en fonction de « l'orientation sociale » d'un agent :

Catégorie d'orientation sociale	Valeur de δ
Agresseur	$\delta < 0$ et très important en valeur absolue
Compétiteur	$\delta < 0$ et relativement important en valeur absolue
Individualiste	δ proche de zéro
Coopérateur	$\delta > 0$ et relativement important
Altruiste	$\delta > 0$ et très important

TABLE 5.2.: Les valeurs prises par δ en fonction de « l'orientation sociale » d'un individu

des traits de personnalité (représenté ici par la valeur des paramètres) que i est en mesure de déceler chez son partenaire.

L'évaluation du comportement d'autrui

La définition formelle de θ_{ij} implique que, lors de la prise en compte d'autrui, la tendance *inconditionnelle* du joueur i à valoriser les gains de son partenaire, voit son influence agir conjointement avec celle d'un second terme qui évalue la bienveillance de l'agent j à l'égard de l'agent i .

Ce second terme est défini comme étant égal à :

$$\varepsilon_{ij} = \alpha_i(y_j - \bar{y}_i) \quad (5.2.5)$$

Où α_i représente la sensibilité de l'agent i à l'évaluation du comportement de son partenaire.

Plus précisément, nous supposons, à travers cette spécification, que chaque agent $i \in I$ évalue la bienveillance de son partenaire j ($i \neq j$) sur la base de la façon dont l'action de j peut être comparée à la norme de contribution internalisée par i . En outre, nous pensons que cette évaluation conduit à la détermination de l'appréciation *conditionnelle* que l'agent i développe, en ce qui concerne l'issue du jeu pour l'agent j . Finalement, à la différence de Bowles et Gintis (2005), qui omettent le rôle des émotions lors de l'évaluation de la bienveillance d'un partenaire, nous supposons que cette appréciation *conditionnelle* se produit de concert avec l'émergence d'émotions particulières et que cela fait naître chez l'agent un désir d'expression réciproque.

Ainsi, nous considérons le rôle de deux émotions discrètes pouvant intervenir lors de cette évaluation.

En premier lieu, si la contribution de l'autre agent est trop faible, tel que $y_j < \bar{y}_i$, ε_{ij} est négatif. Dans un tel cas, l'agent i évalue l'agent j comme malveillant. Nous interprétons cette analyse comme conjointe à (voire résultante de) l'influence de l'émotion de colère. Celle-ci se déclenche lorsqu'une norme est violée, ce qui est le cas quand le comportement de j ne respecte pas la norme de contribution de i . Nous considérons que l'émergence de cette émotion participe à la prise de conscience, par i , que le comportement de j n'est pas à la hauteur de ce que i considère comme moralement juste de contribuer.

En second lieu, si nous supposons au contraire que $\varepsilon_{ij} > 0$, la gratitude est alors l'émotion positive qui se manifeste et encourage l'agent i à considérer j comme bienveillant à son égard. En effet, la gratitude émerge lorsque l'on profite d'une action coûteuse et intentionnelle, de la part d'un autre agent, sans aucune compensation directe. Nous considérons ici que i compte sur j pour faire correspondre son comportement avec ce que i a intériorisé comme étant la norme de contribution. Ainsi, lorsque j entend contribuer plus que ce que i attend, dans la mesure où, d'une part, l'agent i bénéficie de l'investissement de j au bien public et, d'autre part, l'acte de contribuer est coûteux (en comparaison avec le fait de conserver son argent pour soi), cela conduit à l'émergence de sentiments de gratitude chez i . A nouveau, nous soutenons que l'intervention de l'émotion de gratitude aide l'agent i à ouvrir les yeux sur la bienveillance relative dont l'agent j fait preuve lorsque son comportement répond aux attentes du premier.

5. Le Cas Statique

Au total, dès lors que $\alpha_i \neq 0$, on peut considérer que le joueur i , est un coopérateur conditionnel.

Nous supposons, pour chaque agent $i \in I$, que l'influence conjointe de la valorisation *inconditionnelle* des gains de j et de l'évaluation *conditionnelle* de la bienveillance de ce dernier, se mue finalement en θ_{ij} , le poids global accordé aux gains de j dans le jeu. Cette prise en compte du sort de j pouvant s'effectuer de manière positive ou négative, selon le signe de θ_{ij} .

Si l'on suppose que $\delta_i \geq 0$ pour tout $i \in I$. Il existe deux cas pour lesquels θ_{ij} peut être positif :

- D'abord, quand $\varepsilon_{ij} > 0$, θ_{ij} est à l'évidence positif. En effet, cela implique à la fois que i éprouve de la sympathie et un sentiment de proximité par rapport à l'agent j , mais aussi qu'il se sent reconnaissant envers lui, en raison du niveau de contribution que celui-ci met en œuvre. L'une des principales tendances à l'action de la gratitude est de créer un besoin de faire preuve de réciprocité positive envers un bienfaiteur. Ici, du fait que $\theta_{ij} > 0$, l'agent i valorise positivement les gains de j et, en conséquence, obtient de la satisfaction lorsqu'ils augmentent. Ainsi, ceci motive l'agent i à entreprendre des actions qui bénéficient à son partenaire. Dans la version standard du JBP, la seule option dont dispose l'agent i afin d'agir en faveur de j , est d'augmenter son niveau de contribution au bien public. Ce faisant, i exprime des signes de réciprocité positive. De toute évidence, à mesure que θ_{ij} augmente, i est de mieux en mieux disposé à l'égard de j et, par conséquent, sent une plus grande envie d'agir prosocialement envers lui.
- L'autre possibilité, pour que θ_{ij} soit positif, est que l'on ait $\varepsilon_{ij} < 0$, mais que, dans le même temps, $|\varepsilon_{ij}| < \delta_i$. Avant d'aller plus loin, il convient de se rappeler que, comme nous avons pu le voir précédemment, la proximité ressentie par un individu, vis-à-vis d'un autre agent, est en corrélation avec sa capacité à pardonner cet agent. De plus, nous supposons ici que cette corrélation s'étend également à sa capacité à faire preuve d'indulgence à l'égard de ce dernier. Cette capacité est intégrée à notre formalisation à travers δ_i (lorsque celui-ci est positif), qui représente une force de défense face aux émotions négatives ressenties par i à l'égard de j , lorsque le comportement de ce dernier n'est pas en accord avec les principes moraux de i . Ainsi, plus δ_i est élevé, plus il est difficile pour i d'être dominé par la colère. Dans le cas présent, le capital de sympathie dont j dispose auprès de i est entamé mais n'est pas épuisé, et il permet toujours à i d'accorder un poids positif au sort de j dans le jeu, bien que ce dernier n'agisse pas suffisamment correctement pour mériter légitimement une telle considération. Cependant, le manque de sentiments de gratitude implique une faible valeur de θ_{ij} et, par conséquent, une motivation moindre, pour l'agent i , à agir de manière prosociale.

Parallèlement, toujours dans le cas où l'on suppose $\delta_i \geq 0$, pour que θ_{ij} soit négatif, cela nécessite que $\varepsilon_{ij} < 0$ et que $|\varepsilon_{ij}| > \delta_i$. Dans ces conditions, le poids négatif accordé finalement aux gains du joueur j est interprété comme la prévalence du sentiment de colère chez l'agent i . Dans un tel cas, l'ampleur de l'agression que l'agent i perçoit dans le com-

portement de l'agent j épuise le capital de sympathie dont j dispose auprès de i , lequel n'offre alors qu'une résistance partielle aux sentiments négatifs de i . L'une des principales tendances à l'action de la colère est la recherche de vengeance, possiblement à travers l'utilisation d'actes de punition. Or, de manière cohérente avec cette tendance à l'action, l'évaluation négative du sort de j implique que i dérive de l'utilité de la diminution des gains de son partenaire, ce qui est une force motivant la mise en œuvre d'actions de rétorsion. Ainsi, cela se traduit par une diminution du consentement à coopérer chez l'agent i , réduire sa contribution est en effet le seul moyen à travers lequel celui-ci peut exprimer son désir de réciprocité négative, dans la version standard du JBP.

Enfin, il convient de noter que, pour tout agent $i \in I$, $i \neq j$, il est possible que $|\theta_{ij}| > 1$. Formellement, cela signifie que le poids que l'agent i accorde aux gains matériels de j surpasse, en valeur absolue, le poids que celui-ci accorde à ses propres gains dans le jeu. Nous justifions cette conception en supposant que ce qui influence l'utilité des agents n'est pas tant lié aux simples gains matériels qu'ils obtiennent dans le jeu, mais davantage aux conséquences émotionnelles associées au fait de parvenir à ces issues matérielles. Ainsi, puisque les humains sont des êtres sensibles et sociaux, il est tout à fait crédible de penser que, lors de la participation à une situation de dilemme social, telle que le JBP, l'utilité émotionnelle liée à la façon dont l'interaction se déroule avec un partenaire (représentée ici par $\theta_{ij}\pi_j$) puisse dépasser, en terme d'intensité, le plaisir égocentrique provenant de l'obtention de gains matériels dans le jeu (représenté par π_i).

Dans la perspective des éléments de simulation que nous allons proposer dans la suite du document, il convient, avant de passer à la résolution du modèle, de présenter certaines hypothèses concernant la valeur que peuvent prendre les paramètres composant la fonction d'utilité.

5.2.1.1.3. Hypothèses sur les paramètres

La norme de contribution au bien public

Dans notre modélisation, s'agissant de la détermination, pour chaque joueur i , de la norme de contribution qu'il utilise dans le jeu du bien public, il convient de noter que la valeur de cette norme est la même en ce qui concerne l'évaluation qu'il effectue de son propre comportement ou de celui de son partenaire.

Ce choix est somme toute en adéquation avec certains travaux en psychologie dont il ressort que le comportement mis en œuvre par un individu est souvent utilisé par celui-ci comme un repère permettant de juger de la validité du comportement des autres (Alicke, Dunning et Krueger, 2005). A la différence qu'ici, le point de référence n'est pas un comportement effectif mais une valeur préféfinie, antérieurement intériorisée.

Par ailleurs, il est souvent admis, dans le cas du jeu du bien public, que la moyenne des contributions est un repère dans l'évaluation du comportement d'un partenaire dans un groupe. Par exemple, Fehr et Fischbacher (2004b) expliquent que dans l'une de leurs expériences, les individus dont le niveau de contribution était inférieur au niveau moyen ont été punis, tandis que ceux dont le niveau de contribution était supérieur au niveau moyen ne l'ont presque pas été.

5. Le Cas Statique

De notre côté, nous choisissons un processus « hybride » pour la détermination de la norme de contribution de chaque agent. Le terme hybride est utilisé ici dans le sens où chaque joueur a sa propre définition de la valeur de la norme, ce qui donne de l'importance aux convictions de l'agent (ce qui se rapproche d'une évaluation de l'adéquation des comportements sur la base même de son propre comportement), mais que, dans le même temps, sa norme de référence correspond à sa perception personnelle du comportement moyen des individus, au sein de l'économie, dans cette situation (ce qui se rapproche d'une définition basée sur un comportement moyen).

L'avantage d'une telle définition est qu'elle semble être plus pertinente s'agissant de la prise en compte du sentiment de culpabilité. En effet, si la référence morale, dans une quelconque situation, se limite au comportement que l'on met en œuvre, est-il possible de se sentir coupable ? En outre, lorsque le nombre de joueurs est égal à deux, une norme de comportement égale à la moyenne des contributions implique qu'il n'existe aucun cas où chaque joueur est à la fois satisfait de son propre comportement et du comportement de son partenaire (hormis lorsqu'ils se comportent tous les deux de la même façon). Or, nous pensons que lorsque deux individus contribuent de manière conséquente (bien que différente) à un bien public, il est possible que ces deux individus soient satisfaits d'avoir contribué suffisamment et que l'individu qui a contribué le plus ne soit pas dérangé par la contribution légèrement plus faible de son partenaire. Notre spécification permet de tenir compte de ce type de situation, dans le cas où les contributions des deux joueurs dépassent le niveau de la plus exigeante des normes de contribution.

Le capital de sympathie δ

Nous avons vu que le paramètre δ représente le niveau de sympathie et de proximité qu'un agent ressent envers son partenaire dans le jeu. Dans la majorité des expériences qui sont conduites dans le cadre du jeu du bien public, les sujets ne se connaissent pas, la participation s'effectuant généralement de manière anonyme (les joueurs prenant leurs décisions sur un ordinateur, en étant isolés les uns des autres). Dans ces conditions, étant donné qu'il ne s'est produite aucune interaction concrète entre les partenaires avant le début du jeu, on peut considérer que le paramètre δ est relatif à « l'orientation sociale » de l'individu.

Empiriquement, il se dégage que les individus sont, en majorité, des individualistes ou des coopérateurs (Liebrand, 1984 ; Offerman, Sonnemans et Schram, 1996). Par exemple, dans une des expériences menées par Liebrand (1984), 10 % des sujets sont qualifiés de compétitifs, 31 % sont qualifiés d'individualistes, 53 % sont qualifiés de coopératifs et 5 % sont qualifiés d'altruistes (les chiffres peuvent cependant varier d'une expérience à l'autre).

De plus, les résultats du test de l'anneau, réalisé dans l'expérience de Sonnemans *et al.* (2006), montrent qu'avoir une « orientation sociale » négative (caractérisant les agresseurs et les compétiteurs) ne se produit presque jamais.

Par conséquent, nous supposons que pour tout agent $i \in I$, $\delta_i \geq 0$.

En outre, les équations (5.2.3) et (5.2.4) impliquent que si le comportement du partenaire de i respecte exactement la valeur de la norme de contribution de ce dernier, soit $y_j = \bar{y}_i$, alors $\theta_{ij} = \delta_i$. Par conséquent, dans une telle situation, les poids que i accorde à

ses propres gains matériels dans le jeu et à ceux du joueur j , sont respectivement égaux à 1 et δ_i . Etant donné que, dans une telle situation, i se trouve dans un état émotionnel neutre vis-à-vis de son partenaire, suite à l'analyse du comportement de celui-ci (puisque le comportement de j correspond exactement à la valeur de la norme de i), il semble assez naturel de considérer que i accorde, au plus, le même poids au sort de son partenaire qu'à son propre sort à l'issue du jeu. Cette argumentation est appuyée par la faible représentativité observée des agents ayant une « orientation sociale » altruiste.

Par conséquent, on admet également que $\delta_i \leq 1, \forall i \in I$.

Les paramètres de sensibilité à l'égard du respect de la norme

Les deux paramètres α_i et β_i sont tous deux caractéristiques de la sensibilité et de l'importance que le joueur i accorde au respect de la norme de comportement \bar{y}_i . Cependant, α_i désigne la sensibilité de l'agent i au fait que la contribution d'autrui respecte \bar{y}_i , tandis que β_i représente la sensibilité que l'agent i démontre en ce qui concerne son propre respect de la norme de comportement qu'il a internalisée.

Le fait de ressentir des sentiments (positifs ou négatifs), en conséquence du comportement adopté par autrui, participe sans doute (par des processus de généralisation) à notre compréhension de la façon dont autrui est à même d'interpréter notre propre comportement et donc à la détermination de ce qui peut être attendu de nous, dans un contexte particulier. Un tel phénomène fait écho à la maxime de la règle d'or : « Ne faites pas aux autres ce que vous ne voudriez pas que l'on vous fasse », qui formule un procédé, retrouvé de façon universelle, quant à la prise de conscience de la moralité d'un acte. Ainsi, notre sensibilité au fait qu'autrui respecte un certain comportement normatif, semble nous fournir une information indicative de l'importance que nous accordons à la norme en question et, par extension, à celle que la société devrait lui accorder, de notre point de vue.

Par conséquent, lorsqu'un individu est sensible au respect, de la part des autres, d'une certaine norme de comportement, celui-ci est conscient du fait que les autres sont probablement également sensibles au fait que la norme soit respectée (y compris par lui), et que la moralité de ses conduites est prise en compte dans l'évaluation qu'autrui fait de lui. Au demeurant, d'après Wojciszke (2005), lors de l'évaluation d'une personne, l'examen de la moralité de sa conduite importe plus que l'examen de ses compétences. A cet égard, savoir adopter des comportements appropriés sur le plan moral est à nouveau un atout crucial pour être accepté des autres et développer des relations avec eux.

Pour le bon fonctionnement des sociétés humaines, l'éducation, ainsi que de nombreux autres processus de socialisation, ont été développés, de sorte à inculquer un goût pour le fait d'agir moralement chez les individus. Certaines personnes sont ainsi capables d'agir moralement, même en l'absence d'incitations économiques ou stratégiques (punitions, récompenses), uniquement du fait de leur désir de respecter les principes qu'ils ont identifiés comme étant moraux. De telles motivations à respecter, quelles que soient les circonstances, les normes identifiées comme moralement pertinentes dans la société, sont désignées par Batson (2008) comme étant caractéristiques de l'intégrité morale d'un individu. Selon nous, la culpabilité est une émotion qui a été sélectionnée afin de favoriser l'intégrité morale des individus.

5. Le Cas Statique

En ce sens, dans le cadre de notre modèle, on suppose que, pour tout $i \in I$, la valeur de α_i (qui est indicative de l'ampleur avec laquelle l'agent considère la norme de contribution comme pertinente, puisqu'elle caractérise l'intensité avec laquelle une violation de la norme est susceptible de le heurter) n'est pas trop éloignée de celle de β_i (représentative de la sensibilité de l'agent au fait de respecter la norme qu'il a identifiée). Ainsi, si la norme de contribution au bien public est importante aux yeux de l'agent i , puisqu'il est sensible au fait que les autres la respectent (c'est-à-dire que celui-ci est caractérisé par une forte valeur de α_i) et qu'en parallèle, celui-ci fait preuve de cohérence dans son évaluation morale des comportements, alors il ne peut pas être insensible à la question de savoir si son propre comportement est conforme à la valeur de la norme (ce qui implique une valeur de β_i qui ne soit pas trop faible par rapport à celle de α_i). De même, si l'agent i est faiblement sensible au fait qu'autrui respecte ou non la norme de contribution qu'il a internalisée (ce qui peut être indicatif de la faible pertinence de cette norme, selon lui), il serait contre intuitif de penser que celui-ci soit très concerné par le fait de respecter cette norme, à travers son propre comportement.

Pour autant, des recherches en psychologie font état de nombreux individus ne démontrant pas un tel niveau de cohérence dans leurs jugements moraux. En effet, une large partie des agents font preuve d'hypocrisie morale, qui consiste à être animé par le souhait d'apparaître comme une personne morale (pour des considérations de réputation et d'acceptation au sein des groupes), tout en évitant, lorsque c'est possible, le coût lié au fait d'agir moralement (Batson et Thompson, 2001; Batson, Collins, et Powell, 2006; Batson, 2008). Ainsi, d'aucun pourrait objecter que l'hypothèse selon laquelle α_i et β_i ne peuvent pas avoir des valeurs trop différentes est remise en question par de tels résultats. Cependant, dans le contexte particulier du jeu du bien public, le comportement de chaque agent est clairement identifiable par l'ensemble des partenaires dans le jeu (du moins dans la version du jeu que nous analysons). En conséquence, le seul moyen pour les individus d'apparaître comme des agents moraux, est de se comporter en adéquation avec leur norme morale. De fait, même si l'on ne peut négliger la présence d'agents faisant preuve d'hypocrisie morale, au sein de l'économie, on peut ignorer ce phénomène dans le cadre de notre modèle. Par conséquent, nous supposons que les agents qui sont affectés (colère, déception) lorsque la norme de contribution est violée, dans une certaine mesure, par autrui, ressentent des sentiments négatifs (culpabilité sincère, ou seulement embarras de se retrouver à découvert) d'intensité comparable, lorsqu'ils sont eux-mêmes à l'origine d'une violation de la norme, d'ampleur équivalente.³

Par conséquent, on admet que pour tout $i \in I$, $\beta_i \leq \alpha_i < 2\beta_i$ ⁴.

De plus, par souci de simplicité, on suppose que $\beta_i \in]0, 1]$ ⁵ et que $\alpha_i \leq 1$.

3. Il pourrait être envisageable qu'un agent i hypocrite sur le plan moral puisse, contrairement à un individu moralement intègre, être capable de souhaiter apparaître comme quelqu'un ayant des aptitudes morales (forte valeur de β_i), lorsque son comportement est observé, même dans le cas où le respect de cette norme, par les autres, n'est pas primordial pour lui (faible valeur de α_i). Cependant, nous choisissons d'ignorer cette possibilité puisque nous pensons qu'il est relativement rare de faire face à une personne qui accorde de l'importance au fait de respecter une norme, même en étant hypocrite, sans avoir un minimum d'exigence vis-à-vis du comportement d'autrui.

4. La valeur 2 est arbitraire et a été choisie par souci de simplicité et de commodité mathématique.

5. Le cas $\alpha_i = \beta_i = 0$ n'est possible que s'agissant d'un individu de type *homo oeconomicus*, pour lequel tous les autres paramètres du modèle sont également nuls. Dans tous les autres cas, on suppose

5.2.1.2. La résolution du modèle

Dans cette version standard du JBP, chaque agent $i \in I$ maximise son utilité par rapport à son niveau de contribution. Il est important de préciser ici que nous supposons que les agents se trouvent en condition d'information parfaite et complète. Ceux-ci connaissent donc parfaitement les caractéristiques de leur partenaire et sont capables, en conséquence, d'anticiper parfaitement ses réactions. De même, cette hypothèse implique que chaque agent est en mesure de parfaitement anticiper l'émergence des émotions, chez lui comme chez son partenaire, pour toute issue possible du jeu.

La condition de premier ordre est :

$$\frac{\partial U_i}{\partial y_i} = 0$$

$$\Leftrightarrow -1 + a + \theta_{ij}a + 2\beta_i(\bar{y}_i - y_i) = 0$$

Il apparaît alors que cette dérivée est composée de $-1 + a$, le coût marginal lié au fait de contribuer au bien public ; $\theta_{ij}a$, qui désigne l'évaluation par i de l'effet marginal de sa contribution sur les gains de l'autre joueur ; et, finalement, $2\beta_i(\bar{y}_i - y_i)$, la variation marginale du sentiment de culpabilité consécutive à une variation du niveau de contribution. La condition de premier ordre conduit à la définition de la fonction de meilleure réponse suivante :

$$y_i^R = \frac{-1 + a + a\theta_{ij} + 2\beta_i\bar{y}_i}{2\beta_i}$$

Pout tout $i \in I$, $i \neq j$, $y_i^R(y_j)$ définit la meilleure réponse stratégique du joueur i en fonction du niveau de contribution de l'agent j .

Avec $\theta_{ij} = \delta_i + \alpha_i(y_j - \bar{y}_i)$, cette fonction de meilleure réponse s'écrit :

$$y_i^R(y_j) = \frac{-1 + a + a(\delta_i + \alpha_i(y_j - \bar{y}_i)) + 2\beta_i\bar{y}_i}{2\beta_i}$$

$$\Leftrightarrow y_i^R(y_j) = \frac{a\alpha_i}{2\beta_i}y_j + \frac{-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i}{2\beta_i} \quad (5.2.6)$$

$\alpha_i > 0$ et $\beta_i > 0$, $\forall i \in I$.

5. Le Cas Statique

Pour tout $i \in I$, $i \neq j$, cette dernière expression nous permet de mieux nous rendre compte du fait que les fonctions de meilleure réponse des agents sont des fonctions linéaires et croissantes du niveau de contribution consenti par l'autre joueur, ce qui implique géométriquement que ces fonctions sont représentées par des droites dans le plan (y_i, y_j) . Grâce à cette expression, il est également aisé de voir que $\frac{\partial y_i^R}{\partial \delta_i} > 0$, ce qui indique qu'une augmentation de l'intensité du sentiment de sympathie et de proximité de l'agent i envers l'autre joueur conduit, toutes choses égales par ailleurs, à un comportement plus coopératif, par le déplacement vers le haut de sa fonction de meilleure réponse, comme on peut le voir dans le graphique suivant (où les courbes en pointillé représentent les fonctions de meilleure réponse des deux agents après une augmentation, toutes choses égales par ailleurs, du niveau de leurs capital de sympathie) :

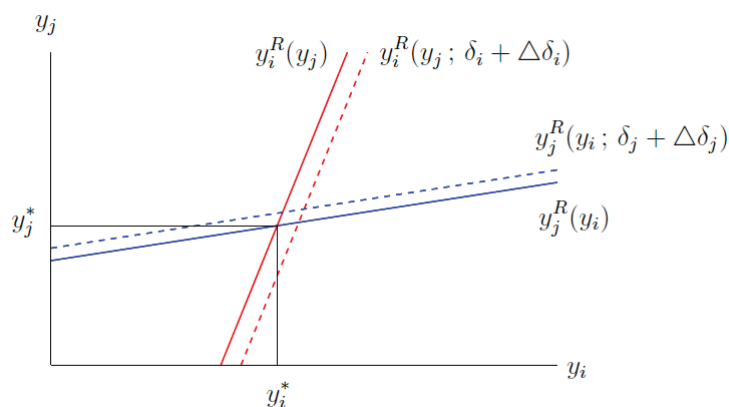


FIGURE 5.2.1.: Fonctions de meilleure réponse

Graphiquement, le point d'intersection entre les deux fonctions de meilleure réponse correspond à l'unique équilibre de Nash du jeu. Analytiquement, cet équilibre est obtenu par l'utilisation d'un raisonnement *à la Cournot* et, pour tout $i \in I$, $i \neq j$, il est égal à :

$$y_i^* = \frac{a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i]}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} \quad (5.2.7)$$

5.2.1.3. Résultats

L'expression de l'équilibre de Nash que nous venons de présenter est relativement lourde. Ainsi, pour mieux défricher l'analyse des résultats de notre modèle, nous nous intéressons, en premier lieu, au cas symétrique, pour lequel, pour tout $i \in I$, $i \neq j$:

$$\alpha_i = \alpha_j = \alpha \quad \beta_i = \beta_j = \beta \quad \delta_i = \delta_j = \delta \quad \text{et} \quad \bar{y}_i = \bar{y}_j = \bar{y}$$

5.2.1.3.1. Le cas symétrique

L'équilibre de Nash du jeu, dans le cas symétrique, devient :

$$y_i^* = y_j^* = y^* = \frac{-1 + a(1 + \delta) + (2\beta - a\alpha)\bar{y}}{(2\beta - a\alpha)}$$

$$\Leftrightarrow y^* = \frac{-1 + a(1 + \delta)}{(2\beta - a\alpha)} + \bar{y} \quad (5.2.8)$$

Il apparaît alors clairement que la norme de contribution commune, que les deux agents ont intériorisée, revêt une importance capitale dans leur prise de décision. Ce résultat est cohérent avec le fait que l'on suppose que la norme de contribution s'interprète, pour chaque agent, comme le comportement stratégique qu'il convient moralement de choisir, par défaut, afin d'être accepté par autrui. De plus, un tel résultat est illustratif de la capacité des normes à orienter le comportement des agents dans les sociétés humaines. En effet, le poids de la loi, des conventions, des traditions, du conformisme est dûment établi dans le domaine des sciences humaines.

Toutefois, on constate également que nos agents sont susceptibles de s'écarter de ce point de référence, sous l'influence de l'incitation économique à coopérer (a), de l'intensité avec laquelle ils ressentent les émotions morales (β et α), ou encore du niveau de sympathie, envers leur partenaire, qui les caractérise (δ).

On remarque également que :

$$y^* = \frac{-1 + a(1 + \delta)}{(2\beta - a\alpha)} + \bar{y} \geq \bar{y}$$

$$\Leftrightarrow \frac{-1 + a(1 + \delta)}{2\beta - a\alpha} \geq 0$$

$$\Leftrightarrow (1 + \delta) \geq \frac{1}{a}$$

$$\Leftrightarrow \delta \geq \frac{1 - a}{a} \tag{5.2.9}$$

Ainsi, pour un taux donné de rendement marginal du bien public, la capacité des joueurs à choisir un niveau de contribution qui dépasse la valeur de leur norme commune, dépend positivement de l'intensité de leur capital de sympathie et de la proximité qu'ils ressentent l'un envers l'autre. Par conséquent, il ressort de notre modèle, qu'en regroupant des agents qui ont une « orientation sociale » suffisamment tournée vers les autres (grande valeur de δ), on peut favoriser la coopération de ces agents dans la version standard du jeu du bien public, au sens de les faire contribuer au delà d'une norme de contribution commune. Ce résultat est cohérent avec les données expérimentales issues du travail de Brekke, Hauge, Lind et Nyborg (2011), qui montrent que des groupes formés d'individus ayant, lors d'une première étape, réalisé un geste altruiste (donner à la croix rouge), ont par la suite démontré des niveaux de coopération (et de maintien de la coopération) plus importants, dans le jeu du bien public, comparativement aux niveaux obtenus par des groupes composés de sujets ayant fait le choix de ne pas donner.

Proposition 1. *Dans le cas symétrique, lorsque les agents ont un capital de sympathie $\delta \geq \frac{1-a}{a}$, leur niveau de contribution, à l'équilibre, est supérieur ou égal à la valeur de la norme commune qu'ils ont intériorisée.*

Au total, la proposition 1 met en lumière le fait que, dans cette version du jeu, pour que les agents mettent en œuvre un comportement coopératif, au sens où celui-ci est, au moins, respectueux de la norme établie dans l'économie, il est nécessaire d'avoir un niveau de sympathie partagé par les joueurs d'autant plus grand que l'action coopérative rapporte peu. De façon plus précise, on constate que moins le fait de contribuer au bien public est intéressant, sur le plan économique, plus les agents doivent éprouver de la sympathie l'un envers l'autre pour pouvoir s'investir suffisamment dans l'effort collectif pour que la norme qu'ils ont intériorisée soit, au moins, respectée. Par conséquent, si l'on sait, au moment de recruter des individus pour travailler en équipe, que l'effort coopératif n'est pas très rentable, bien que le coopération soit socialement souhaitable, il est préférable de se tourner vers des individus ayant une « orientation sociale » suffisamment altruiste (grande valeur de δ). Ainsi, le fait que l'on retrouve souvent des personnes qualifiées de généreuses et d'altruistes, dans les milieux associatifs et particulièrement au sein d'associations bénévoles, tend à illustrer la pertinence et la portée générale de la proposition 1.

Par ailleurs, puisque l'on suppose $\delta \leq 1$, on sait que lorsque l'on pose $a = \frac{1}{2}$, soit lorsque la coopération est rendue la moins rentable possible dans le jeu, on obtient $y^* \leq \bar{y}$.

Proposition 2. *Dans le cas symétrique, lorsque le bien public offre un rendement marginal égal à $a = \frac{1}{2}$, le niveau de contribution des agents, à l'équilibre, ne dépasse jamais la valeur de la norme commune qu'ils ont intériorisée.*

Dans une telle situation, lorsque l'action de coopérer est très peu rentable, on constate qu'il est difficile de stimuler la coopération des individus.

Toutefois, les propositions 1 et 2 suggèrent qu'une augmentation du rendement marginal du bien public, par exemple à travers une subvention au financement du bien public⁶, permet de faciliter la mise en œuvre de la coopération, y compris pour des individus ayant un faible capital de sympathie.

Application numérique

Tout au long de cette section, nous proposons des éléments de simulation permettant d'illustrer le fonctionnement de notre modèle. Cependant, ces éléments ne sont présents qu'à titre indicatif et ne servent pas à la réalisation d'une analyse de sensibilité.

Tout d'abord, il convient ici de rappeler les hypothèses que nous posons concernant les paramètres du modèle :

$$\beta \in]0, 1], \alpha \in]\beta, 2\beta[\cap]0, 1], \delta \in [0, 1] \text{ et } \bar{y} \in [0, d]$$

Pour cette première application numérique, nous supposons que les paramètres caractérisant les deux joueurs sont les suivants :

d	δ	\bar{y}	β	α
10	0.5	5	0.5	0.5

TABLE 5.3.: Exemple de paramétrisation - JBP standard - Cas symétrique

Afin d'illustrer les résultats de notre modèle, nous utilisons un tel profil particulier de joueurs (*i.e.* une telle paramétrisation), puisqu'il correspond en quelques sortes à un profil « moyen » de joueur, dans le sens où chacun des paramètres du modèle prend ici une valeur centrale, relativement à l'ensemble des valeurs possibles.

Par ailleurs, une norme de contribution égale à 5 (soit la moitié de la dotation initiale des agents) correspond à la fois à un point focal, facilement identifiable par les individus, mais aussi à la valeur moyenne des contributions observées dans le jeu du bien public joué en un coup (cf. 4.2).

6. On montre mathématiquement que, dans cette version du jeu, il revient au même de subventionner la contribution individuelle que d'augmenter le rendement individuel marginal du bien public.

5. Le Cas Statique

Nous pouvons alors nous intéresser à deux valeurs différentes du rendement marginal du bien public et observer les différents niveaux de contributions obtenus à l'équilibre :

a	0.5	0.75
y^*	4.67	5.2
	$y^* < \bar{y}$	$y^* > \bar{y}$

TABLE 5.4.: Résultats du JBP standard - Cas symétrique

Les résultats chiffrés que nous obtenons donnent une illustration des propositions 1 et 2. En effet, lorsque $a = \frac{1}{2}$, la contribution optimale des agents ne dépasse pas la valeur de la norme dans l'économie (qui est ici de 5). A l'inverse, lorsque $a = \frac{3}{4}$, la contribution optimale est supérieure à la valeur de norme avec $y^* = 5.2 > \bar{y}$.

Pour analyser la manière dont les paramètres du modèle jouent sur le niveau de coopération, nous passons au calcul des dérivées.

Calcul des dérivées

On rappelle que, dans le cas symétrique :

$$y^* = \frac{-1 + a(1 + \delta)}{2\beta - a\alpha} + \bar{y}$$

Le rendement marginal du bien public a

$$\frac{\partial y^*}{\partial a} = \frac{(1 + \delta)2\beta - \alpha}{(2\beta - a\alpha)^2} > 0$$

Interprétation :

Intuitivement, notre modèle prédit qu'un rendement marginal du bien public plus important implique, toutes choses égales par ailleurs, une contribution des agents plus importante à l'équilibre.

Le capital de sympathie δ

$$\frac{\partial y^*}{\partial \delta} = \frac{a}{2\beta - a\alpha} > 0$$

Interprétation :

Cette dérivée est toujours positive, étant donné que l'on suppose que $\alpha < 2\beta$ et que $0 < a \leq 1$. L'interprétation est ici plutôt naturelle : plus les partenaires ont de la sympathie l'un envers l'autre et plus ceux-ci sont poussés à contribuer au bien public, afin de

promouvoir les gains de leur partenaire dans le jeu (on se souvient que $0 \leq \delta$).

En conséquence, comme nous l'avons déjà mentionné plus haut, notre modèle prédit que le fait de sélectionner des participants au jeu ayant une « orientation sociale » tournée vers autrui, ou encore de stimuler le niveau de proximité entre les agents, par exemple, en les faisant communiquer et se rencontrer avant le début du jeu (Balliet, 2009), permet d'observer une coopération élevée.

Statique comparative :

Par rapport à la paramétrisation donnée dans le tableau 5.3, on suppose, toutes choses égales par ailleurs, que la valeur de δ passe de 0.5 à 1, soit :

	δ	\bar{y}	β	α
Valeurs initiales	0.5	5	0.5	0.5
Valeurs finales	1	5	0.5	0.5

On obtient alors :

a	0.5	0.75
équilibre initial	4.67	5.2
équilibre final	5	5.8
Variation	7.1 %	11.5 %

La norme de contribution \bar{y}

$$\frac{\partial y^*}{\partial \bar{y}} = 1$$

Interprétation :

Dans le cas symétrique, on constate à nouveau, avec le calcul de cette dérivée, que la valeur de la norme partagée par les deux agents est déterminante vis-à-vis de leur niveau de coopération à l'équilibre.

Ainsi, notre modèle prédit que la coopération dans le jeu est favorisée lorsque l'on fait jouer ensemble des individus ayant intériorisé une norme de contribution dont la valeur est importante. Cette intériorisation pouvant résulter de l'éducation reçue, originellement et structurellement, par les agents, ou bien être le résultat d'un stimuli social, comme un rappel ou la description de l'existence et de la pertinence d'une norme, avant la participation au jeu. Notre modèle est donc cohérent avec de nombreux résultats empiriques comme, par exemple, ceux obtenus par Goldstein, Cialdini, et Griskevicius (2008), qui montrent que le fait d'indiquer aux usagers d'une chambre d'hôtel que la majorité des autres clients de cet établissement utilise plusieurs fois une même serviette, influe significativement sur le niveau de coopération des individus, ce qui permet de réaliser des économies d'eau et d'énergie.

La sensibilité à l'émotion de culpabilité β

$$\frac{\partial y^*}{\partial \beta} = \frac{-2\beta[-1 + a(1 + \delta)]}{(2\beta - a\alpha)^2}$$

Le signe de cette dérivée dépend du signe de $-1 + a(1 + \delta)$

Ce qui correspond à la même condition décrite par l'inéquation (5.2.9) permettant de déterminer si la contribution des joueurs dépasse ou non le niveau de la norme commune.

Interprétation :

La dérivée est positive uniquement lorsque la contribution des joueurs, à l'optimum, est inférieure à la valeur de la norme commune. Dans ce cas, la culpabilité a une influence positive sur la coopération, et une augmentation de l'intensité avec laquelle les joueurs la ressentent résulte en une augmentation du niveau de contribution à l'équilibre, de façon à se rapprocher de la contribution normative.

Inversement, la dérivée est négative dans le cas où le niveau de contribution des agents, à l'optimum, est supérieur à la valeur de la norme commune (ce qui nécessite des valeurs de δ et a importantes).

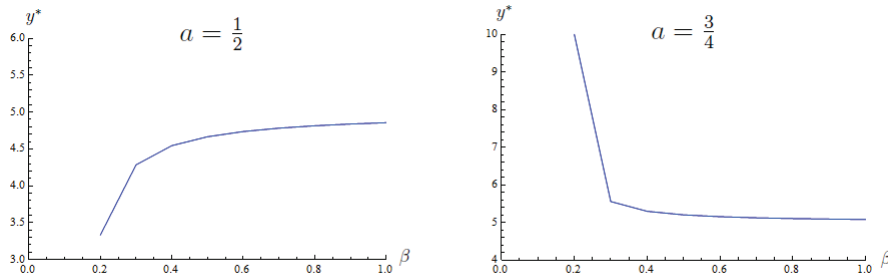
Dans ce cas, une plus forte intensité de la culpabilité implique une contribution plus faible de la part de chacun des joueurs, dans le but de diminuer le coût psychologique lié à une trop forte contribution par rapport à leur norme (du fait de la peur d'être exploité en faisant plus que ce qui est recommandé par la norme, ou du fait de la crainte de se priver de ressources nécessaires au respect d'autres normes de comportements dans les autres dilemmes auxquels ils peuvent faire face). En analysant ce même résultat différemment, on peut voir que, lorsque les individus ont un niveau de contribution supérieur à la valeur de la norme, ce qui implique qu'ils ressentent de la gratitude l'un pour l'autre, alors le fait que ces agents ressentent le sentiment de culpabilité avec une moindre intensité, implique qu'ils ont moins de scrupules à s'éloigner de la contribution normative, afin de laisser s'exprimer leurs sentiments réciproques.

Notre modèle prédit ainsi que lorsqu'il s'agit d'une situation où la coopération est difficile à mettre en œuvre (car elle est peu intéressante économiquement, ou bien parce que les participants ne sont pas très ouverts les uns aux autres), le fait que ceux-ci ressentent intensément l'émotion de culpabilité (du fait d'un trait de leur personnalité, ou bien d'une induction réalisée avant le début du jeu) permet, toutes choses égales par ailleurs, de favoriser la mise en œuvre d'un comportement prosocial et coopératif.

A l'inverse, si le dilemme social est tel qu'un comportement dépassant la valeur de la norme de contribution est facilement obtenu, alors une augmentation, toutes choses égales par ailleurs, de l'intensité avec laquelle l'émotion de culpabilité est ressentie, influe négativement sur le niveau de contribution, puisque dans cette situation, les agents se sentent plus affectés par le fait de se démunir en contribuant plus qu'ils ne devraient.

Statique comparative :

Les deux courbes tracées dans le graphique suivant, représentent respectivement le niveau de contribution optimal y^* obtenu, toutes choses égales par ailleurs, en fonction de la valeur de β , lorsque $a = \frac{1}{2}$ et lorsque $a = \frac{3}{4}$. Les valeurs des autres paramètres du modèle correspondent à ce qui est indiqué dans le tableau 5.3 :

FIGURE 5.2.2.: Variations de y^* en fonction de β

Sur le graphique 5.2.2, on constate bien comment, lorsque $a = \frac{1}{2}$ et que les contributions à l'optimum ne dépassent pas la valeur de la norme, une augmentation de la valeur de β conduit, toutes choses égales par ailleurs, à une augmentation de la valeur des contributions optimales. A l'inverse, lorsque $a = \frac{3}{4}$ et que les contributions optimales dépassent la norme intériorisée par chaque agent, une augmentation de β entraîne, toutes choses égales par ailleurs, une chute drastique du niveau de contribution joué à l'équilibre par les agents.

La sensibilité aux émotions d'évaluation d'autrui α

$$\frac{\partial y^*}{\partial \alpha} = \frac{[-1 + a(1 + \delta)]a}{(2\beta - a\alpha)^2}$$

Le signe de cette dérivée dépend également du signe de $-1 + a(1 + \delta)$, ce qui correspond à nouveau à la condition décrite par l'inéquation (5.2.9).

Interprétation :

Pour chaque agent, le paramètre α représente l'intensité des émotions liées à l'évaluation du comportement de son partenaire. Ainsi, soit il représente la colère, dans le cas où l'agent i est irrité par une trop faible contribution de la part de son partenaire ($y^* \leq \bar{y}$), soit il représente la gratitude, lorsque qu'il perçoit que son partenaire réalise une action qu'il considère comme bienveillante ($y^* \geq \bar{y}$).

La dérivée est positive uniquement dans le cas où la contribution à l'optimum dépasse la valeur de la norme commune (ce qui nécessite des valeurs de δ et a importantes). Lorsque c'est le cas, une augmentation de α , conduit les agents à ressentir la gratitude avec davantage d'intensité et ils sont alors prêts à faire preuve de plus de réciprocité positive en

5. Le Cas Statique

augmentant leur niveau de contribution à l'équilibre. A l'inverse, lorsque la contribution n'est pas suffisamment élevée à l'optimum et qu'elle ne parvient pas à atteindre le niveau de la norme, ce qui signifie que les individus sont sous l'influence de la colère, alors une augmentation de l'intensité avec laquelle cette émotion est ressentie implique la mise en œuvre d'une réciprocité négative d'une ampleur d'autant plus grande, à travers une diminution de la contribution des agents à l'optimum.

Là encore, pour des agents possédant des caractéristiques identiques, notamment en ce qui concerne leur norme de contribution, des profils différents d'agents sont souhaités pour favoriser la coopération, en fonction de la nature du dilemme social en question.

En particulier, lorsque le dilemme social est tel qu'il est difficile d'obtenir un niveau conséquent de coopération, il est préférable, toutes choses égales par ailleurs, d'avoir des agents qui sont moins sensibles à l'évaluation du comportement des autres, parce qu'autrement, ceux-ci sont amenés à ressentir plus intensément l'émotion de colère, ce qui tend à amoindrir le poids que chaque agent accorde au sort de son partenaire et à entraîner une diminution du niveau de coopération.

Au contraire, lorsque la situation est telle que le niveau de la norme de contribution est dépassée, notre modèle suggère qu'il est préférable que des agents relativement plus sensibles au comportement adopté par leur partenaire jouent ensemble. En effet, dans cette situation, étant donné que les agents ressentent l'émotion de gratitude, lorsqu'ils évaluent le comportement de leur partenaire, il est préférable, du point de vue de la promotion de la coopération, que cette émotion soit ressentie avec plus d'intensité.

Statique comparative :

Les deux courbes tracées dans le graphique suivant, représentent respectivement le niveau de contribution optimal y^* obtenu, toutes choses égales par ailleurs, en fonction de la valeur de α , lorsque $a = \frac{1}{2}$ et lorsque $a = \frac{3}{4}$. Les valeurs des autres paramètres du modèle correspondent à ce qui est indiqué dans le tableau 5.3 :

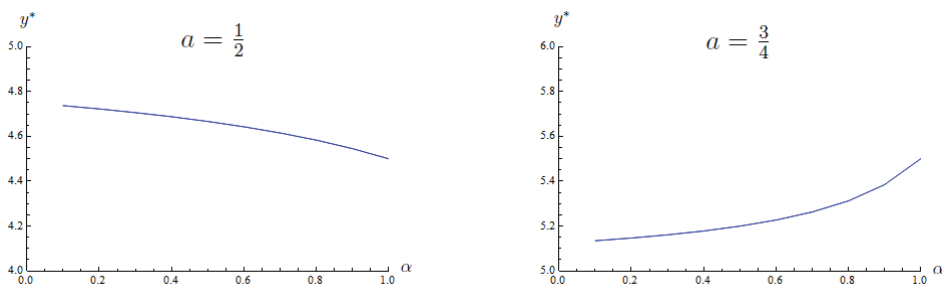


FIGURE 5.2.3.: Variations de y^* en fonction de α

Sur le graphique 5.2.2, on constate bien comment, lorsque $a = \frac{1}{2}$ et que les contributions à l'optimum ne dépassent pas la valeur de la norme, une augmentation de la valeur de α conduit, toutes choses égales par ailleurs, à une diminution de la valeur des contributions

optimales. A l'inverse, lorsque $a = \frac{3}{4}$ et que les contributions optimales dépassent la norme intériorisée par chaque agent, une augmentation de α entraîne, toutes choses égales par ailleurs, une augmentation du niveau de contribution joué à l'équilibre par les agents.

5.2.1.3.2. Le cas hétérogène

Supposons à nouveau que chaque joueur est caractérisé par des valeurs des paramètres qui lui sont propres et qui peuvent être différentes de celles de son partenaire. On rappelle que dans ce cas, pour tout $i \in I, i \neq j$, l'équilibre de Nash obtenu est le suivant :

$$y_i^* = \frac{a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i]}{4\beta_i\beta_j - a^2\alpha_i\alpha_j}$$

Il convient de noter que, puisque l'on suppose que $2\beta_i - a\alpha_i > 0$ et $2\beta_j - a\alpha_j > 0$, il vient :

$$4\beta_i\beta_j - a^2\alpha_i\alpha_j > 0 \tag{5.2.10}$$

Avant de nous intéresser aux calculs des dérivées de la contribution optimale, en fonction des différents paramètres, il convient de mentionner que, dans l'hypothèse où $a = \frac{1}{2}$, la proposition suivante peut être énoncée :

Proposition 3. *Dans le cas hétérogène, lorsque le bien public offre un rendement marginal égal à $a = \frac{1}{2}$, le niveau de contribution des deux agents ne dépasse jamais la valeur de la norme la plus élevée.*

La démonstration de ce résultat est fournie en annexe.

La proposition 3 s'inscrit dans le prolongement des deux précédentes et souligne la difficulté que peut représenter, en termes de niveau de coopération, le fait de jouer au jeu du bien public, dans sa version standard, notamment lorsque celui-ci offre un faible rendement du bien public. En effet, on remarque que malgré le fait que les joueurs aient intériorisé une norme de contribution qui influe sur leur comportement contributif, il est difficile pour eux de se conformer totalement à cette norme à travers leur décision de contribution (et il leur est impossible de pouvoir contribuer au-delà de celle-ci).

Application numérique

Supposons maintenant que les deux joueurs soient caractérisés par les valeurs suivantes des paramètres du modèle :

$i = \{1, 2\}$	δ_i	\bar{y}_i	β_i	α_i
Joueur 1	0.6	5	0.3	0.4
Joueur 2	0.4	4	0.5	0.5

TABLE 5.5.: Exemple de paramétrisation - JBP standard - Cas hétérogène

5. Le Cas Statique

Nous n'attribuons aucun profil de joueur particulier aux valeurs des paramètres que nous retenons pour cet exemple. Cependant, sur la base de notre interprétation des différents paramètres du modèle, nous pouvons décrire la nature des joueurs 1 et 2. Ainsi, le joueur 1 dispose d'un plus grand capital de sympathie que son partenaire et possède également une norme de contribution dont la valeur est légèrement plus importante. Ainsi, on peut s'attendre à ce que le joueur 1 soit plus altruiste que le joueur 2 et qu'il soit ainsi celui dont le niveau de coopération est le plus important à l'équilibre. De son côté, le joueur 2 est celui qui est le plus en prise avec ses émotions, du fait de leurs plus grandes intensités chez lui, comparativement avec celles qui caractérisent le joueur 1.

Lorsque ces deux individus jouent ensemble, les niveaux de contribution que nous obtenons, à l'équilibre, en fonction de différentes valeurs du rendement marginal du bien public, sont les suivants :

$a = 0.5$		
y_1^*	y_2^*	Moyenne du groupe
4.25	3.76	4
$\bar{y}_2 < y_1^* < \bar{y}_1$	$y_2^* < \bar{y}_2 < \bar{y}_1$	

$a = 0.75$		
y_1^*	y_2^*	Moyenne du groupe
5.05	4.44	4.75
$\bar{y}_2 < \bar{y}_1 < y_1^*$	$\bar{y}_2 < y_2^* < \bar{y}_1$	

TABLE 5.6.: Résultats du JBP standard - Cas hétérogène

Les résultats obtenus dans cet exemple montrent comment, dans le JBP standard, le joueur qui a la norme de contribution la plus importante (joueur 1) est en quelque sorte « tiré vers le bas » par son partenaire. En effet, celui-ci ne parvient pas à atteindre la valeur de sa norme de contribution lorsque le rendement marginal du bien public est le plus bas, et ne le dépasse que très légèrement dans le cas où il est le plus favorable ($a = \frac{3}{4}$). Parallèlement, lorsque le rendement marginal du bien public augmente (ce qui peut représenter une subvention des contributions individuelles) la coopération augmente et le joueur 2 contribue plus que la valeur de sa norme, du fait de la plus grande attractivité du comportement coopératif.

Ces résultats peuvent être mis en perspective par rapport à ce qui se passerait si chacun des deux joueurs jouaient avec un partenaire qui lui est identique. Les deux tableaux suivants indiquent respectivement les équilibres obtenus lorsque deux joueurs 1 jouent ensemble et lorsque deux joueurs 2 jouent l'un avec l'autre. On note ces nouveaux équilibres \hat{y}_1 pour le joueur 1 et \hat{y}_2 pour le joueur 2.

\widehat{y}_1	
$a = 0.5$	$a = 0.75$
4.5	5.67
$y_1^* < \widehat{y}_1 < \bar{y}_1$	$\bar{y}_1 < y_1^* < \widehat{y}_1$

\widehat{y}_2	
$a = 0.5$	$a = 0.75$
3.6	4.08
$\widehat{y}_2 < y_2^* < \bar{y}_2$	$\bar{y}_2 < \widehat{y}_2 < y_2^*$

Ici, on remarque bien l'effet que les joueurs 1 et 2 provoquent l'un sur l'autre, lorsqu'ils jouent ensemble. En effet, on voit que le fait de jouer avec le joueur 1 amène le joueur 2 à être plus coopératif, comparativement au cas où il devrait interagir avec un individu ayant les mêmes caractéristiques que lui. Inversement, le joueur 1 contribue moins à l'équilibre en jouant avec 2, que s'il jouait avec un individu qui lui est parfaitement semblable. Ce dernier résultat est illustratif de la capacité de notre modèle à représenter les comportements de coopération conditionnelle, puisque, par exemple, le joueur 1 est ici prêt à être plus coopératif avec quelqu'un qui réagit de la même façon que lui (un autre joueur de son type), plutôt qu'avec quelqu'un de plus égoïste (le joueur 2).

Pour déterminer quels paramètres du modèle jouent, sans ambiguïté, sur le niveau de coopération, nous nous intéressons maintenant au calcul des dérivées.

Calcul des dérivées

Pour chaque paramètre individuel du modèle, l'équilibre de Nash joué par un agent est affecté à la fois par la valeur que prend, chez lui, le paramètre en question, mais aussi par la valeur, prise par ce même paramètre, chez son partenaire. En conséquence, pour chaque paramètre individuel traité ici, on s'intéresse au calcul de la dérivée de la contribution optimale d'un agent, d'une part en fonction de son propre paramètre, et d'autre part en fonction du paramètre de son partenaire.

Le capital de sympathie δ

$$\frac{\partial y_i^*}{\partial \delta_i} = \frac{a2\beta_j}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} > 0 \text{ et } \frac{\partial y_i^*}{\partial \delta_j} = \frac{a^2\alpha_i}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} > 0$$

Interprétation :

L'interprétation est à nouveau assez naturelle ici, plus l'individu i se sent proche de son partenaire, plus il a envie de contribuer afin d'augmenter les gains de celui-ci.

En outre, plus l'individu i est conscient que son partenaire se sent proche de lui, plus sa contribution est importante à l'équilibre. En effet, le fait que le partenaire de l'agent i ressent plus de sympathie à son égard implique que le premier réagit plus favorablement à un même niveau de contribution de la part de ce dernier, ce qui incite alors l'agent i à augmenter son niveau de contribution, afin de bénéficier du fait que la mise en place d'un acte de réciprocité positive est facilitée chez son partenaire.

5. Le Cas Statique

On constate à nouveau tout l'intérêt de recruter des participants ayant une tendance accrue à ressentir de la sympathie à l'encontre des autres, ou encore de susciter le sentiment de sympathie, chez l'un, l'autre ou les deux partenaires, avant la participation au jeu, dans la perspective de promouvoir la coopération dans la version standard du JPB.

La norme de contribution \bar{y}

$$\frac{\partial y_i^*}{\partial \bar{y}_i} = \frac{2\beta_j(2\beta_i - a\alpha_i)}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} \quad \text{et} \quad \frac{\partial y_i^*}{\partial \bar{y}_j} = \frac{a\alpha_i(2\beta_j - a\alpha_j)}{4\beta_i\beta_j - a^2\alpha_i\alpha_j}$$

Avec $0 < \frac{\partial y_i^*}{\partial \bar{y}_i} < 1$ et $0 < \frac{\partial y_i^*}{\partial \bar{y}_j} < 1$.

Interprétation :

Dans le cas hétérogène, la valeur de la norme intériorisée par l'agent i revêt à nouveau une importance capitale dans la modulation du choix stratégique de ce dernier. En effet, le niveau de contribution à l'équilibre de l'agent i dépend positivement du niveau de sa propre conception de la norme, cependant, comparativement au cas symétrique, le comportement de l'agent i est moins sensible face à une variation de celle-ci. Ce résultat est logique puisque, dans le cas symétrique, la norme augmentait pour l'ensemble des deux agents, tandis qu'ici seule la norme de joueur i varie.

De plus, la valeur de la norme intériorisée par le partenaire de l'agent i a également une influence positive sur le comportement du premier. En effet, une augmentation de la norme de l'agent j implique une augmentation de son niveau d'exigence vis-à-vis du comportement de contribution de l'agent i , et cela incite donc l'agent i à augmenter son niveau de contribution à l'équilibre pour continuer à contenter son partenaire, ou pour limiter son mécontentement.

Notre modèle prédit donc qu'il est possible, dans le cas d'agents hétérogènes, de favoriser l'émergence de comportements coopératifs dans le JBP standard, en recrutant au moins un participant ayant une forte norme de contribution, ou en induisant une plus grande valeur de cette norme chez l'un, l'autre, ou les deux partenaires. Ce résultat est cohérent avec les travaux expérimentaux qui mettent en évidence l'influence des leaders sur le comportement de leurs partenaires, dans les groupes de participants au jeu du bien public (cf. 4.2.2.2).

S'agissant des autres paramètres, la complexité de l'interaction entre deux partenaires hétérogènes est telle qu'elle ne permet pas de pouvoir réaliser une interprétation précise des dérivées calculées. Néanmoins, sur la base de nos simulations, il semble que la variation des niveaux de contribution, à l'équilibre, induite par la variation des paramètres de culpabilité et de sensibilité au comportement d'autrui, sont toujours sensibles à la manière dont les contributions, à l'équilibre initial, se comparent à la valeur des normes intériorisées par les deux agents.

5.2.2. Le jeu du bien public avec phase de récompense

Nous considérons maintenant que $n = 2$ joueurs participent au JBP, avec phase subséquente de récompense, que nous avons déjà décrit dans la première section de ce chapitre (cf. 5.1.1.2).

5.2.2.1. La fonction d'utilité

L'ajout d'une phase de récompense implique l'introduction de nouveaux termes dans la fonction d'utilité des agents. Pour clarifier l'exposé, on peut diviser en deux la fonction d'utilité représentant le processus de décision des agents dans cette version du jeu. Pour tout $i \in I$, $i \neq j$, on a alors :

$$U_i = U_i^C + U_i^R \quad (5.2.11)$$

Avec

$$U_i^C = AC_i + EA_{ij}$$

Où U_i^C se réfère alors à l'utilité liée principalement à la phase de contribution, que l'on a déjà présentée ci-dessus (équation (5.2.2)).

Puisque la phase de contribution reste la même par rapport à la version standard du JBP, il n'est pas nécessaire de présenter à nouveau la fonction U_i^C , nous passons donc directement à la présentation de la fonction U_i^R :

$$U_i^R = R_{ij} + R_{ji} \quad (5.2.12)$$

Où U_i^R désigne l'utilité qui se rapporte principalement à la phase de récompense ⁷.

Plus précisément, la partie R_{ij} de la fonction U_i^R se réfère à l'utilité que l'agent i retire de l'acte de récompenser son partenaire, avec :

$$R_{ij} = \theta_{ij}r_{ij} - \beta_i r_{ij}^2 - cr_{ij} \quad (5.2.13)$$

7. Cette notation est quelque peu abusive car chaque joueur i est supposé utiliser ses gains π_i , obtenus lors la phase de contribution, pour réaliser une éventuelle récompense. Ainsi, en définissant U_i^R , comme nous l'avons fait, on néglige cet aspect, en raison de l'absence de π_i dans son expression. Néanmoins, cette présentation n'a pour but que d'alléger l'exposé et π_i est bien présent dans la fonction d'utilité globale de chaque joueur i .

5. Le Cas Statique

Où $\theta_{ij} = \delta_i + \alpha_i(y_j - \bar{y}_i)$ et où r_{ij} est le montant de récompense reçu par l'agent j , de la part de l'agent i .

Parallèlement, la partie R_{ji} de cette même fonction U_i^R se rapporte à l'utilité que l'agent i dérive du fait d'être récompensé par son partenaire, avec :

$$R_{ji} = (1 + \rho_i)r_{ji} \quad (5.2.14)$$

Où r_{ji} est le montant de récompense reçu par l'agent i , de la part de l'agent j , et ρ_i est un paramètre individuel présenté ci-après.

Ainsi, on peut également écrire la fonction d'utilité de tout agent $i \in I$, $i \neq j$, sous la forme suivante :

$$U_i = \pi_i - \beta_i(\bar{y}_i - y_i)^2 + \theta_{ij}\pi_j + \theta_{ij}r_{ij} - \beta_i r_{ij}^2 - cr_{ij} + (1 + \rho_i)r_{ji} \quad (5.2.15)$$

Où les gains obtenus par les joueurs durant la phase de contribution π_i et π_j sont toujours définis par l'équation (5.1.1).

Comme dans le cas de la version standard, si on pose $\alpha_i = \delta_i = \beta_i = \rho_i = 0$, on retrouve la représentation classique du processus de décision des agents, en théorie des jeux.

Pour poursuivre la présentation de notre fonction d'utilité dans le détail, nous allons d'abord nous intéresser à l'utilité liée à l'acte de récompenser (fonction R_{ij}), avant de passer ensuite à la description de l'utilité qui est relative au fait d'être récompensé (fonction R_{ji}).

5.2.2.1.1. L'utilité liée à l'acte de récompenser

Un nouveau moyen d'expression des sentiments réciproques

A travers le premier terme de la fonction R_{ij} , qui est $\theta_{ij}r_{ij}$, on perçoit tout l'esprit du fonctionnement du modèle, dans la version du JBP avec phase de récompense. En effet, pour tout $i \in I$, $i \neq j$, θ_{ij} représente toujours le poids accordé, par i , au sort de j dans le jeu, sur la base de l'influence conjointe de la valorisation *inconditionnelle* des gains de j (le capital de sympathie) et de l'évaluation *conditionnelle* de la bienveillance de ce dernier, laquelle dépend de l'action stratégique réalisée par j lors de la phase de contribution. De fait, on remarque que ce poids accordé aux gains de j , déterminé sur la base de l'action menée par celui-ci, durant la phase de contribution, s'applique également, lors de la phase de récompense, de sorte à définir la façon dont le joueur i valorise le fait de récompenser son partenaire, en fonction de la contribution de ce dernier, lors de la phase initiale.

Ainsi, on peut remarquer que, lorsque $\theta_{ij} > 0$, ce qui implique que j a adopté un comportement suffisamment approprié vis-à-vis de la norme de son partenaire, $\theta_{ij}r_{ij}$ est interprété comme un gain psychologique issu du simple fait de récompenser l'autre agent. On retrouve ici l'influence de la gratitude, dont la principale tendance à l'action est de conduire à la mise en place d'une réciprocité positive entre les agents. Par conséquent, lorsque j a agi suffisamment correctement, selon i , lors de la phase de contribution, on suppose que i valorise le fait de le récompenser et trouve un certain plaisir à le faire. Ce plaisir peut également se rapporter à ce qu'Andreoni (1989) a qualifié de « goût de donner ». La récompense offerte par i bénéficie directement à j , on peut donc observer que, comparativement à la version standard du jeu du bien public, l'introduction de cette seconde phase, dans le déroulement du jeu, offre un nouveau moyen d'expression à la gratitude du joueur i (que le simple fait de contribuer de manière conséquente, lorsque j en fait de même), qui est considérablement plus ciblé et donc plus facilement interprétable par j .

A l'inverse, lorsque $\theta_{ij} < 0$, ce qui signifie que le joueur j ne parvient pas à se conformer aux attentes de i lors de la phase de contribution, $\theta_{ij}r_{ij}$ représente le coût psychologique associé au fait de récompenser j . Ce coût se rapporte alors à l'influence de la colère, dont la tendance à l'action est la recherche de vengeance (soit une forme de réciprocité négative). En conséquence, lorsque j transgresse la norme de contribution de son partenaire par son comportement lors de la première phase, on suppose que i valorise négativement le fait de le récompenser et donc que cet acte est déplaisant pour lui. Dans ce cas de figure, l'agent i est à nouveau en mesure, au cours de la deuxième étape du jeu, d'exprimer son sentiment (d'insatisfaction ici) à l'égard du comportement du joueur j par son niveau de « récompense ». Néanmoins, le mieux qu'il puisse faire pour cela, dans cette version du jeu, se résume à ne pas récompenser du tout son partenaire.

Le coût lié à l'acte de récompenser son partenaire

Le terme, $-cr_{ij}$, représente le coût monétaire que l'agent i doit supporter pour offrir à j une récompense d'un montant r_{ij} . Ceci est une différence clé par rapport au modèle présenté par Bowles et Gintis (2005). En effet, les auteurs ont basé leur formalisation sur les protocoles expérimentaux utilisés dans le travail séminal de Fehr et Gächter (2000), lesquels sont caractérisés par une fonction de coût (supporté pour punir un autre agent) qui est non linéaire. Une telle non linéarité est opportune, sur le plan mathématique, en ce sens qu'elle est nécessaire pour permettre le calcul d'un niveau optimal de punition qui soit différent des solutions en coin (je ne dépense rien pour punir mon partenaire, ou bien je dépense tout ce que j'ai pour le punir). Ainsi, un possible revers de leur formalisation pourrait être le fait que celle-ci dépende de cette forme spécifique de la fonction de coût, pour fonctionner d'une manière réaliste, alors même que, dans la littérature, il existe de nombreux protocoles expérimentaux dans lesquels le coût à supporter pour infliger une punition est une fonction linéaire de la quantité de sanction choisie.

Dans notre spécification, nous introduisons plutôt un autre terme, $-\beta ir_{ij}^2$, qui représente un coût psychologique lié à l'acte de récompenser l'autre joueur. La présence de cette perte supplémentaire d'utilité est justifiée de la même façon que pour le coût psychologique engendré, chez un joueur lorsqu'il contribue au-delà de sa norme (déjà analysé ci-dessus). Nous supposons donc que les agents peuvent se sentir coupables de consacrer (trop) de ressources dans le but de remercier leur partenaire, alors que ces ressources

pourraient être nécessaires afin de se conformer à l'une des autres normes de comportement qu'ils ont intériorisées et qui concernent d'autres dilemmes ou situations.

En parallèle, chaque agent pourrait également être averse au fait de se placer en position de vulnérabilité lorsqu'il récompense l'autre joueur, puisqu'à l'inverse de l'investissement qu'il réalise en contribuant au bien public, les ressources dépensées pour récompenser sont définitivement perdues pour l'agent (il n'y a pas de retour sur investissement).

5.2.2.1.2. L'utilité liée au fait d'être récompensé ou la satisfaction d'être reconnu pour son comportement

Finalement, $R_{ji} = (1 + \rho_i)r_{ji}$ mesure l'utilité dérivée par l'agent i , du fait d'être récompensé par l'agent j . Ce gain est à la fois d'ordre matériel et psychologique, puisque l'on suppose que $\rho_i r_{ji}$ est la part psychologique du gain obtenu par i , lorsque celui-ci est récompensé d'un montant r_{ji} par son partenaire. En effet, le niveau de récompense reçu par l'agent i peut être envisagé comme un indicateur l'informant du niveau d'approbation sociale que souhaite lui communiquer son partenaire. C'est également une manière, pour un partenaire, de communiquer la gratitude que ce dernier ressent en réponse de l'effort coopératif de l'agent i .

Un exemple concret illustrant ce besoin d'être rassuré à propos de l'appréciation de l'effort coopératif que l'on fournit se retrouve dans la pratique des sports collectifs. Une telle pratique nécessite un degré de coopération très important pour qu'une équipe soit performante dans la compétition et ce qui est observé, quasi universellement et de manière presque mécanique et constante, ce sont les gestes et signes d'appréciation et de reconnaissance que les joueurs d'une même équipe développent entre eux. Tous se tapent dans la main à chaque bonne action, se congratulent, parfois à l'excès, dès qu'une action décisive a été accomplie et ce manège est continu durant l'intégralité d'un match. Notre interprétation de ces signes de cohésion d'équipe est que la pratique du sport, surtout à haut niveau, est un exercice très difficile et que la communication de l'approbation est un geste qui permet de « booster » la confiance d'un individu, par un renforcement de l'estime de soi, et que cela est nécessaire pour réussir à un tel niveau de difficulté. En ce sens, il semble réaliste de considérer que ce gain psychologique découle, au moins en partie, de l'émergence de l'émotion de fierté, qui intervient à mesure que l'agent récompensé est témoin de la valorisation sociale positive qui est attribuée à son comportement. Ainsi, le paramètre ρ_i mesure l'intensité avec laquelle l'émotion de fierté est ressentie, à la marge, par l'agent i , lorsqu'il est récompensé.

A ce stade, il est important de discerner dans notre formulation, une autre distinction existant entre notre modèle et celui sur lequel nous appuyons. En effet, on se souvient que dans le modèle de Bowles et Gintis (2005), l'impact des émotions qui émergent chez l'agent i , lorsqu'il est sanctionné, diffère selon la manière dont son propre comportement se conforme à la norme qu'il a intériorisée. Ainsi, l'agent souffre du sentiment de honte uniquement lorsqu'il est sanctionné *et* que sa contribution est inférieure au niveau de la norme. A l'inverse, lorsque $y_i > \bar{y}_i$, le terme introduit par les auteurs représente l'influence de la rancune, provenant de la contrariété que l'agent i ressent du fait d'être puni alors même que, selon ses propres standards de comportement, il agit de manière appropriée.

Dans notre formalisation, l'influence des émotions lorsque l'on est récompensé est toujours positive. Lorsqu'un agent est récompensé pour avoir adopté un comportement qui dépasse ses standards de contribution, il semble naturel qu'il en ressente le sentiment de fierté. Qu'en est-il lorsque sa contribution est en deçà de la valeur de sa norme ? Nous pensons que chaque agent est tout de même rassuré de recevoir un tel signe de reconnaissance, de la part de son partenaire, et donc, de se rendre compte que son comportement est valorisé par autrui. Nous pensons ainsi que, dans un tel cas, l'émotion ressentie par l'agent récompensé, bien qu'elle puisse être différente du sentiment de fierté, a une valence positive et qu'elle implique un gain d'utilité.

Par ailleurs, nous pensons que si nous avons construit notre forme fonctionnelle de manière symétrique par rapport au modèle de Bowles et Gintis (2005), vis-à-vis de ce point particulier, cela aurait signifié qu'un agent puisse se sentir doublement coupable de la réalisation d'un comportement qu'il considère comme inapproprié. En effet, il se sentirait coupable, d'une part au moment de contribuer, dans le cas où il ne parvient pas à respecter sa norme et, d'autre part, lorsqu'il est récompensé pour son comportement, alors qu'il n'est pas parvenu à se conformer à sa norme de contribution. En réalité, ce type de phénomène, où la culpabilité initiale est accentuée par une approbation que l'on juge iméritée ne peut être exclue. Cependant, il nous semble que cette nuance soit de l'ordre du détail et que l'on puisse intégrer l'influence du sentiment de culpabilité à travers un seul terme, tout en considérant que, même chez les personnes les plus exigeantes envers elles-mêmes, recevoir l'approbation de la part d'autrui puisse malgré tout être un facteur d'atténuation de ce sentiment.

Enfin, chez Bowles et Gintis (2005) l'insertion d'une telle nuance est importante car c'est une façon d'introduire le risque d'escalade des sentiments négatifs qui peut avoir lieu entre partenaires (Hopfensitz et Reuben, 2009). Dans leur modèle, la honte doit être ressentie pour que le message de désapprobation influence le comportement de l'agent de sorte à le faire agir de manière plus coopérative et, pour cela, celui-ci doit se voir communiquer des émotions négatives, de la part de son partenaire (à travers la punition), uniquement lorsqu'il se considère lui-même comme fautif. Dans le cas contraire, la rancune intervient et entretient un climat hostile entre les deux agents.

Une telle possibilité de propagation des émotions est possible dans le cadre de la procédure d'incitation que constitue l'opportunité de récompenser, bien que cela s'effectue de manière plus positive et constructive. Cette fois-ci, le signal envoyé par le partenaire, à travers l'acte de récompense, qui communique une émotion positive (la gratitude), a tendance à transmettre ce ressenti positif et à faire naître en retour une émotion positive chez l'agent qui reçoit la récompense, peu importe si celui-ci se sent coupable ou non de ne pas avoir contribué plus. Cet élément nous permet donc d'expliquer en quoi le fait de cultiver une communication positive entre les agents peut les conduire à entretenir des relations cordiales, là où une communication qui serait orientée vers les aspects négatifs de l'interaction comprend un risque de dégradation de ces relations.

5.2.2.1.3. Hypothèses sur les paramètres

Par souci de simplicité et d'unification, on suppose que le paramètre ρ_i , mesurant la sensibilité émotionnelle de l'individu i au fait de recevoir l'approbation de son partenaire, à travers le niveau de récompense de celui-ci, est tel que $\rho_i \in [0, 1]$.

Il convient également de remarquer que le paramètre β_i intervient à la fois lors de la décision de contribution et lors du choix du niveau de récompense. Ceci nous semble pertinent puisque, dans ces deux situations, le rôle de ce paramètre est d'introduire une aversion des individus au fait d'adopter un comportement qui n'est pas conforme à la norme en vigueur et/ou qui menace de les démunir d'un minimum de ressource.

La différence la plus notable entre ces deux situations étant que notre spécification implique une norme de récompense nulle ($\bar{r}_i = 0$) au sein de l'économie. Ce choix nous semble réaliste, puisque l'acte de récompense est un acte conditionnel par essence. Il n'y a donc pas de raison que les agents aient intégré une norme concernant un montant de récompense à mettre inconditionnellement en œuvre dans une telle situation. A l'inverse, s'agissant du financement d'un bien public, on imagine plus volontier que les individus aient pu intérioriser une norme de comportement de contribution qu'ils considèrent comme juste, même s'ils sont par ailleurs animés par des considérations réciproques face au comportement de leurs partenaires.

5.2.2.2. La résolution du jeu

Le jeu se résout en recourant à un raisonnement par induction à rebours. Dans un premier temps, on détermine, pour chaque agent $i \in I$, le niveau optimal de récompense, qui dépend du niveau de contribution mis en place par le partenaire, ce qui conduit à définir une fonction de récompense. Ensuite, en supposant que chaque agent utilise cette fonction de récompense, on peut à nouveau définir des fonctions de meilleure réponse pour chacun des joueurs, en termes de niveau de contribution, lesquelles permettent la détermination d'un unique équilibre de Nash sur les niveaux de contribution.

La condition de premier ordre donne :

$$\frac{\partial U_i}{\partial r_{ij}} = \theta_{ij} - 2\beta_i r_{ij} - c = 0$$

$$\Leftrightarrow 2\beta_i r_{ij} = \theta_{ij} - c$$

$$r_{ij}^* = \begin{cases} \frac{\delta_i + \alpha_i(y_j - \bar{y}_i) - c}{2\beta_i} & \text{si } \delta_i + \alpha_i(y_j - \bar{y}_i) - c > 0 \\ 0 & \text{sinon} \end{cases} \quad (5.2.16)$$

Il convient de remarquer que $\delta_i + \alpha_i(y_j - \bar{y}_i) - c > 0 \Leftrightarrow y_j > \bar{y}_i + \frac{c - \delta_i}{\alpha_i}$.

Ainsi, on note $\underline{y}_j = \bar{y}_i + \frac{c - \delta_i}{\alpha_i}$, le niveau de contribution qui doit être dépassé, par le joueur j , à l'équilibre, pour que le joueur i lui accorde une récompense positive.

Nous pouvons observer que le niveau de récompense accordé dépend positivement de la sympathie que le joueur i a vis-à-vis de son partenaire et, surtout, du niveau de contribution consenti par celui-ci. Une telle relation de proportionalité, qui semble tout à fait naturelle, est pertinente avec des données empiriques sur les comportements de récompense, dans le jeu du bien public, comme celles dégagées dans le travail de Choi et Ahn (2013). Nous supposons que chaque agent utilise cette fonction optimale de récompense dans le jeu.

Par conséquent, nous pouvons résoudre le modèle pour le niveau de contribution.

La condition de premier ordre conduit à la définition de la fonction de meilleure réponse suivante :

$$y_i^R = \frac{a\alpha_i}{2\beta_i}y_j + \frac{-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i + (1 + \rho_i)\frac{\alpha_j}{2\beta_j}}{2\beta_i} \quad (5.2.17)$$

En supposant une solution intérieure, l'unique équilibre de Nash du JBP avec phase de récompense est :

$$y_{r_i}^* = \frac{a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j + (1 + \rho_j)\frac{\alpha_i}{2\beta_i}] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i + (1 + \rho_i)\frac{\alpha_j}{2\beta_j}]}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} \quad (5.2.18)$$

Il convient de noter que, pour que cet équilibre soit mis en œuvre, on doit avoir $y_{r_i}^* > \underline{y}_j$ pour tout $i \in I$, $i \neq j$. Dans le cas contraire, si aucune de ces conditions n'est remplie, les joueurs savent qu'à l'équilibre, personne ne récompense. Ainsi, dans un tel cas, l'introduction de la phase de récompense ne peut pas être en mesure d'influer sur les joueurs et la résolution du jeu aboutit au même équilibre de Nash que celui qui a été trouvé dans la version standard du JBP (équation (5.2.7)).

5. Le Cas Statique

5.2.2.3. Résultats

Premièrement, lorsque l'équilibre est implémenté, on montre que le niveau de contribution obtenu dans le JPB avec phase de récompense, est strictement supérieur à celui qui est obtenu dans le cas du JBP standard.

En effet, on pose $\forall(i, j) \in I^2, i \neq j, \alpha_i > 0, \beta_i > 0$ et $\rho_i \geq 0$.

Ceci implique que $\forall(i, j) \in I^2, i \neq j, (1 + \rho_j) \frac{\alpha_i}{2\beta_i} > 0$ et $(1 + \rho_i) \frac{\alpha_j}{2\beta_j} > 0$.

Par conséquent, d'après les équations (5.2.7) et (5.2.18), on en déduit que, pour tout $i \in I$:

$$y_{r_i}^* > y_i^*$$

Proposition 4. *Le niveau de contribution joué à l'équilibre par les agents, dans le jeu du bien public avec phase de récompense, est strictement supérieur à celui qui est joué dans le jeu du bien public standard.*

Comme dans le JBP standard, l'expression de l'équilibre de Nash obtenu dans le JPB avec phase de récompense est lourde et pour mieux analyser les résultats de notre modèle, nous nous intéressons, maintenant, au cas symétrique, pour lequel, pour tout $i \in I, i \neq j$:

$$\alpha_i = \alpha_j = \alpha \quad \beta_i = \beta_j = \beta \quad \delta_i = \delta_j = \delta \quad \text{et} \quad \bar{y}_i = \bar{y}_j = \bar{y} \quad \text{et} \quad \rho_i = \rho_j = 0$$

5.2.2.3.1. Le cas symétrique

L'équilibre de Nash du jeu, dans le cas symétrique, devient :

$$y_{r_i}^* = y_{r_j}^* = y_r^* = \frac{-1 + a(1 + \delta) + (2\beta - a\alpha)\bar{y} + (1 + \rho) \frac{\alpha}{2\beta}}{2\beta - a\alpha}$$

$$\Leftrightarrow y_r^* = \frac{-1 + a(1 + \delta) + (1 + \rho) \frac{\alpha}{2\beta}}{2\beta - a\alpha} + \bar{y} \tag{5.2.19}$$

A nouveau, il apparaît clairement que la norme de contribution des agents, conditionne leur prise de décision dans le JBP avec récompense, en raison du fait qu'elle correspond, pour chacun d'eux, au comportement stratégique qui doit être choisi, par défaut, pour être accepté d'autrui.

Toutefois, on constate également que nos agents sont encore davantage susceptibles de s'écarter de ce point de référence, dans le JBP avec récompense, du fait de l'incitation économique et émotionnelle (d'intensité ρ) supplémentaire à coopérer qui est liée à la possibilité d'être récompensé.

On remarque également que :

$$y_r^* = \frac{-1 + a(1 + \delta) + (1 + \rho)\frac{\alpha}{2\beta}}{2\beta - a\alpha} + \bar{y} \geq \bar{y}$$

$$\Leftrightarrow -1 + a(1 + \delta) + (1 + \rho)\frac{\alpha}{2\beta} \geq 0 \quad (5.2.20)$$

A nouveau, la capacité des joueurs à contribuer au-delà de la norme commune dépend positivement de l'intensité de leur capital de sympathie et du rendement marginal du bien public. Toutefois, comparativement au cas standard, celle-ci est également une fonction positive du montant supplémentaire de récompense obtenu pour chaque unité investie dans le bien public $\frac{\alpha}{2\beta}$, ainsi que de l'intensité avec laquelle l'émotion de fierté est ressentie, à la marge, par les agents lorsqu'ils sont récompensés. Ainsi, par rapport au JPB standard, la condition qui doit être remplie pour que les joueurs contribuent plus que la valeur de leur norme est moins exigeante dans le cas du JBP avec récompense.

Par conséquent, il ressort de notre modèle qu'il est beaucoup plus aisé d'obtenir un niveau de coopération au moins égal à la norme de contribution en vigueur dans l'économie, dans le cas du JBP avec récompense que dans le cas du JBP standard.

Application numérique

Supposons que $c = 0.5$ ⁸ et que les paramètres caractérisant les deux joueurs soient les suivantes :

δ	\bar{y}	β	α	ρ
0.5	5	0.5	0.5	0.5

TABLE 5.7.: Exemple de paramétrisation - JBP Récompense - Cas symétrique

8. Utiliser une valeur différente du coût forfaitaire à récompenser c ne modifie pas la nature des résultats.

5. Le Cas Statique

Nous pouvons alors nous intéresser à deux valeurs différentes du rendement marginal du bien public et observer les différents niveaux de contributions obtenus à l'équilibre :

	a	0.5	0.75
JBP standard	y^*	4.67	5.2
JBP récompense	y_r^*	5.67	6.4
	r^*	0.33	0.7
		$y_r^* > \bar{y}$	$y_r^* > \bar{y}$

TABLE 5.8.: Résultat du JBP récompense - Cas symétrique

Les résultats chiffrés que nous obtenons ici, illustrent bien la proposition 4. En effet, que ce soit lorsque $a = \frac{1}{2}$, ou lorsque $a = \frac{3}{4}$, la contribution optimale des agents, dans le JBP avec récompense, est supérieure à celle qui était obtenue dans le cas du JBP standard. De plus, dans chacun des cas, la contribution des agents, à l'équilibre, dépasse la valeur de la norme dans l'économie (qui est ici de 5). Enfin, on remarque également que les joueurs utilisent l'option de récompense, alors que celle-ci est coûteuse.

Calcul des dérivées

On rappelle que, dans le cas symétrique :

$$y_r^* = \frac{-1 + a(1 + \delta) + (1 + \rho)\frac{\alpha}{2\beta}}{2\beta - a\alpha} + \bar{y}$$

Le rendement marginal du bien public a

$$\frac{\partial y_r^*}{\partial a} = \frac{(1 + \delta)2\beta + (1 + \rho)\frac{\alpha^2}{2\beta} - \alpha}{(2\beta - a\alpha)^2} > 0$$

Interprétation :

Comme dans le cas du JBP, notre modèle prédit qu'un rendement marginal du bien public plus important implique, toutes choses égales par ailleurs, une contribution des agents plus importante à l'équilibre, dans le JBP avec récompense.

Le capital de sympathie δ et norme de contribution \bar{y}

Les dérivées du niveau de contribution optimal des agents par rapport à ces deux paramètres sont les mêmes que dans le cas du JBP standard, car δ et \bar{y} n'apparaissent pas dans l'expression du coefficient directeur de la fonction de récompense.

Concernant le paramètre δ , l'interprétation est à nouveau assez naturelle, plus les individus ont de la sympathie les uns pour les autres, plus ils ont envie de contribuer afin d'augmenter les gains de leur partenaire.

Par ailleurs, en ce qui concerne \bar{y} , il semble relativement réaliste que l'influence, à la marge, de la norme de contribution se fasse indépendamment de l'introduction d'une opportunité de récompense. En effet, le fait que cette option soit disponible pour les agents, ne modifie en rien la façon dont la norme sert de point de référence à l'évaluation du comportement de contribution des agents.

Notre modèle prédit donc que la coopération dans le JBP avec récompense est favorisée lorsque les individus qui jouent ensemble ont intériorisé une norme de contribution dont la valeur est importante, ainsi que lorsqu'ils se sentent proches les uns des autres.

La sensibilité à l'émotion de culpabilité β

$$\frac{\partial y_r^*}{\partial \beta} = \frac{-[(1 + \rho)\frac{\alpha}{2\beta^2}(2\beta - a\alpha) + 2(-1 + a(1 + \delta) + (1 + \rho)\frac{\alpha}{2\beta})]}{(2\beta - a\alpha)^2}$$

Le signe de cette dérivée dépend du signe de $-1 + a(1 + \delta) + (1 + \rho)\frac{\alpha}{2\beta}(2 - \frac{a\alpha}{2\beta})$

Ce qui se rapproche de la condition décrite par l'inéquation (5.2.20), permettant de déterminer si la contribution des joueurs dépasse ou non le niveau de la norme commune.

Interprétation :

Etant donné que $2 - \frac{a\alpha}{2\beta} > 1$ (car on a supposé $2\beta > \alpha \geq a\alpha$, cf. 5.2.1.1.3), on constate que, dans le JBP avec récompense, la condition pour que $\frac{\partial y_r^*}{\partial \beta}$ soit négative est moins exigeante que dans le JBP standard, ce qui est cohérent avec le fait que la fonction de récompense dépend négativement de β . Autrement dit, une augmentation, toutes choses égales par ailleurs, de la valeur de β implique un montant de récompense moins important, ce qui peut contribuer à diminuer l'incitation à coopérer, de sorte à limiter le coût psychologique lié au fait de se démunir en contribuant trop et en récompensant trop son partenaire. En analysant ce même résultat sous un autre angle, on peut s'apercevoir que, lorsque les individus ont un niveau de contribution supérieur à la valeur de la norme (impliquant un niveau de récompense mutuelle conséquent, notamment en raison de la gratitude qu'ils ressentent l'un pour l'autre), alors une diminution de l'intensité de leur sentiment de culpabilité leur permet d'exprimer leurs sentiments réciproques avec encore plus de force.

5. Le Cas Statique

Toutefois, lorsque le niveau de coopération à l'équilibre est bas, $\frac{\partial y_r^*}{\partial \beta}$ peut être positif et, dans ces conditions, une augmentation, toutes choses égales par ailleurs, du niveau d'intensité avec lequel les individus se sentent coupables permet une augmentation des contributions à l'équilibre.

Ainsi, notre modèle prédit à nouveau que lorsque la coopération est difficile à susciter, le fait que les agents ressentent intensément l'émotion de culpabilité permet, toutes choses égales par ailleurs, de favoriser la mise en œuvre d'un comportement prosocial et coopératif.

A l'inverse, si le dilemme social est tel qu'un comportement dépassant la valeur de la norme de contribution est facilement obtenu, alors une augmentation, toutes choses égales par ailleurs, de l'intensité avec laquelle l'émotion de culpabilité est ressentie, influe négativement sur le niveau de contribution, puisque dans cette situation, les agents se sentent plus affectés par le fait de se démunir en contribuant plus qu'ils ne devraient et en récompensant plus.

Toutefois, on a vu que l'ajout d'une phase de récompense agit en faveur de la coopération entre les agents. Il semble donc que dans cette situation, une stimulation du niveau de pression face à la norme ne soit pas la bienvenue, car elle serait susceptible de réduire le niveau de contribution des agents.

Statique comparative :

Par rapport à la paramétrisation donnée dans le tableau 5.5, on suppose, toutes choses égales par ailleurs, que la valeur de β passe de 0.5 à 1, ainsi :

	δ	\bar{y}	β	α	ρ
Valeurs initiales	0.5	5	0.5	0.5	0.5
Valeurs finales	0.5	5	1	0.5	0.5

On obtient alors :

situation initiale	a	0.5	0.75
	y_r^*	5.67	6.4
	r^*	0.33	0.7

situation finale	a	0.5	0.75
	y_r^*	5.1	5.31
	r^*	0	0.1

La sensibilité aux émotions d'évaluation d'autrui α

$$\frac{\partial y_r^*}{\partial \alpha} = \frac{(1 + \rho) + (-1 + a(1 + \delta))a}{(2\beta - a\alpha)^2}$$

Le signe de cette dérivée dépend du signe de $-1 + a(1 + \delta) + (1 + \rho)\frac{1}{a}$, une condition qui est proche de celle décrite par l'inéquation (5.2.20).

Interprétation :

Etant donné que $\frac{1}{a} > \frac{\alpha}{2\beta}$ (car on a supposé $2\beta > \alpha \geq a\alpha$, cf. 5.2.1.1.3), alors la condition pour que $\frac{\partial y_r^*}{\partial \beta}$ soit positive est moins exigeante dans le JBP avec récompense que dans le JBP standard. Dans le cas uniforme du JBP avec phase de récompense, on montre ainsi que, si la contribution des agents, à l'équilibre, *dépasse la valeur de la norme de contribution*, une augmentation de leur sensibilité à l'évaluation du comportement d'autrui (sous l'influence de sentiments de gratitude plus intenses, et donc de niveaux de récompense plus importants) une augmentation des contributions à l'optimum. Toutefois, un niveau de contribution des agents inférieur à la valeur de la norme n'implique pas forcément ici que $\frac{\partial y_r^*}{\partial \alpha} \leq 0$.

Lorsque $\frac{\partial y_r^*}{\partial \alpha} \leq 0$, le fait que les contributions ne soient pas suffisamment élevées, à l'équilibre, implique que les individus sont sous l'influence de la colère et donc qu'une augmentation de l'intensité avec laquelle cette émotion est ressentie résulte en une diminution du niveau de coopération à l'optimum.

Ainsi, on montre à nouveau que, lorsque le dilemme social est tel qu'il est difficile d'obtenir un niveau conséquent de coopération, il est préférable, toutes choses égales par ailleurs, que les agents soient moins sensibles à l'évaluation du comportement de leur partenaire, pour ressentir moins intensément l'émotion de colère, de sorte à réhausser le poids que chaque agent accorde au sort de son partenaire et à entraîner un plus grand niveau de coopération.

Au contraire, lorsque la coopération est facilement obtenue, avec des niveaux de contribution proches de la valeur de la norme, notre modèle suggère qu'il est préférable que des agents sensibles au comportement adopté par leur partenaire jouent ensemble. Dans une telle situation, les agents sont en effet amenés à ressentir de la gratitude, plutôt que de la colère, lorsqu'ils évaluent le comportement de leur partenaire, et le niveau de coopération est donc plus important dès lors que cette émotion est ressentie avec plus d'intensité.

Compte tenu du fait que le niveau de coopération est plus important dans le JBP avec récompense, notre modèle suggère que ce mécanisme incitatif est d'autant plus efficace lorsque les individus sont attentifs aux comportements de leur partenaire.

Statique comparative :

Par rapport à la paramétrisation donnée dans le tableau 5.5, on suppose, toutes choses égales par ailleurs, que la valeur de α passe de 0.5 à 1, ainsi :

	δ	\bar{y}	β	α	ρ
Valeurs initiales	0.5	5	0.5	0.5	0.5
Valeurs finales	0.5	5	0.5	1	0.5

On obtient alors :

	a	0.5	0.75
situation initiale	y_r^*	5.67	6
	r^*	0.33	0.5

	a	0.5	0.75
situation finale	y_r^*	7.5	10
	r^*	2.5	6.5

La sensibilité à l'émotion de fierté ρ

$$\frac{\partial y_r^*}{\partial \rho} = \frac{\alpha}{2\beta(2\beta - a\alpha)} > 0$$

Interprétation :

Assez logiquement, le montant de contribution des agents dépend positivement de l'intensité avec laquelle le sentiment de fierté est ressenti par les joueurs, lorsqu'ils sont récompensés. Ainsi, si les joueurs sont plus sensibles, toutes choses égales par ailleurs, au fait de recevoir l'approbation de leur partenaire, sous la forme d'une récompense, ceux-ci sont incités à plus contribuer, de façon à recevoir un montant de récompense plus important.

Par conséquent, notre modèle suggère que l'influence bénéfique du mécanisme d'incitation de récompense est intensifiée lorsque le besoin des individus d'être reconnu pour leurs actions prosociales est accru. Une politique (éducative) visant à développer, chez les agents, un goût à recevoir l'approbation sociale semble ainsi être un outil permettant de promouvoir la coopération, notamment dans le cas où les individus ont l'opportunité de se congratuler.

Statique comparative :

Par rapport à la paramétrisation donnée dans le tableau 5.5, on suppose, toutes choses égales par ailleurs, que la valeur de ρ passe de 0.5 à 1, ainsi :

	δ	\bar{y}	β	α	ρ
Valeurs initiales	0.5	5	0.5	0.5	0.5
Valeurs finales	0.5	5	0.5	0.5	1

On obtient alors :

	a	0.5	0.75
situation initiale	y_r^*	5.67	6
	r^*	0.33	0.5

	a	0.5	0.75
situation finale	y_r^*	6	6.8
	r^*	0.5	0.9

5.2.2.3.2. Le cas hétérogène

Supposons à nouveau que chaque joueur possède ses propres valeurs des paramètres. Pour tout $i \in I, i \neq j$, l'équilibre de Nash obtenu du JBP avec récompense, dans le cas hétérogène :

$$y_{r_i}^* = \frac{a\alpha_i[-1 + a(1 + \delta_j)] + (2\beta_j - a\alpha_j)\bar{y}_j + (1 + \rho_j)\frac{\alpha_j}{2\beta_j} + 2\beta_j[-1 + a(1 + \delta_i)] + (2\beta_i - a\alpha_i)\bar{y}_i + (1 + \rho_i)\frac{\alpha_i}{2\beta_i}}{4\beta_i\beta_j - a^2\alpha_i\alpha_j}$$

Application numérique

Supposons que le coût forfaitaire à récompenser est toujours égal à $c = 0.5$ et que les deux joueurs sont caractérisés par les valeurs suivantes des paramètres du modèle :

$i = \{1, 2\}$	δ_i	\bar{y}_i	β_i	α_i	ρ_i
Joueur 1	0.6	5	0.3	0.4	0.6
Joueur 2	0.4	4	0.5	0.5	0.5

TABLE 5.9.: Exemple de paramétrisation - JBP récompense - Cas hétérogène

Il convient de remarquer ici que l'on retrouve les deux mêmes joueurs 1 et 2 que lors de l'analyse du JBP standard (cf. 5.2.1.3.2). Cependant, l'introduction de la phase de récompense nécessite de fixer la valeur de ρ . Ainsi, ce que l'on peut ajouter à propos de ces deux joueurs et que le joueur 1 est plus sensible que le joueur 2 au fait de recevoir l'approbation sociale qui lui est communiquée, par son partenaire, à travers le niveau

5. Le Cas Statique

de récompense de ce dernier, sous l'effet d'un sentiment de fierté plus intensément ressenti.

Nous pouvons alors nous intéresser à deux valeurs différentes du rendement marginal du bien public et observer les différents niveaux de contributions obtenus à l'équilibre (afin de mieux comparer le JBP avec récompense et le JBP standard, les contributions des joueurs dans la version standard du jeu sont ajoutées entre parenthèse) :

$a = 0.5$		
$y_{r_1}^*$	$y_{r_2}^*$	Moyenne du groupe
6.55 (4.25)	5.67 (3.76)	6.11 (4)
$r_{12}^* = 0.62$	$r_{21}^* = 1.47$	
$\bar{y}_2 < \bar{y}_1 < y_{r_1}^*$	$\bar{y}_2 < \bar{y}_1 < y_{r_2}^*$	

$a = 0.75$		
$y_{r_1}^*$	$y_{r_2}^*$	Moyenne du groupe
8.12 (5.05)	7.24 (4.44)	7.68 (4.75)
$r_{12}^* = 1.66$	$r_{21}^* = 2.45$	
$\bar{y}_2 < \bar{y}_1 < y_{r_1}^*$	$\bar{y}_2 < \bar{y}_1 < y_{r_2}^*$	

TABLE 5.10.: Résultats du JBP récompense - Cas hétérogène

Comparativement à l'application numérique faite dans le cas du JBP standard, en utilisant les mêmes valeurs des paramètres pour chacun des joueurs, on observe, dans le JBP avec récompense, une augmentation des contributions réalisées à l'équilibre. On note également que, dans cet exemple, chaque joueur contribue un montant qui dépasse largement la valeur de la norme de contribution la plus importante (celle du joueur 1), quel que soit le rendement marginal offert par le bien public.

Cette augmentation des contributions est permise par l'introduction de l'opportunité de récompense, laquelle est utilisée par chacun des joueurs ici. Cette option offre aux joueurs un moyen de communiquer avec leur partenaire pour lui exprimer leur approbation, ainsi que leurs sentiments de gratitude, en réponse à l'effort coopératif consenti.

Cet exemple illustre donc bien la capacité du mécanisme incitatif de récompense dans la promotion de la coopération. En outre, les résultats de notre modèle suggèrent que le fait de combiner l'opportunité de récompense, dans le JBP, avec une subvention individuelle des contributions (ce qui correspond à une augmentation du rendement marginal du bien public), permet de susciter encore plus de coopération.

Ces résultats peuvent, comme dans le cas du jeu standard, être mis en perspective avec ce qui se passerait si chacun des deux joueurs jouaient avec un partenaire qui lui est identique. Les deux tableaux suivants indiquent respectivement les équilibres qui seraient obtenus si deux joueurs 1 jouaient ensemble et si deux joueurs 2 jouaient l'un avec l'autre. On note ces nouveaux équilibres \hat{y}_{r_1} pour le joueur 1 et \hat{y}_{r_2} pour le joueur 2. Par ailleurs, on note également \hat{r}_1 et \hat{r}_2 les niveaux de récompense offerts respectivement par les joueurs 1 et 2, à un partenaire qui leur est identique.

Joueurs 1	
$a = 0.5$	$a = 0.75$
$\hat{y}_{r_1} = 7.16$	$\hat{y}_{r_1} = 9.22$
$\hat{r}_1 = 1.61$	$\hat{r}_1 = 2.98$
$\bar{y}_1 < y_1^* < \hat{y}_1$	$\bar{y}_1 < y_1^* < \hat{y}_1$

Joueurs 2	
$a = 0.5$	$a = 0.75$
$\hat{y}_{r_2} = 4.6$	$\hat{y}_{r_2} = 5.28$
$\hat{r}_2 = 0.2$	$\hat{r}_2 = 0.54$
$\bar{y}_2 < \hat{y}_2 < y_2^*$	$\bar{y}_2 < \hat{y}_2 < y_2^*$

Ici, l'effet que les joueurs 1 et 2 ont l'un sur l'autre, lorsqu'ils interagissent ensemble est criant. En particulier, lorsque le rendement marginal du bien public est le plus grand, là où un groupe de deux joueurs 1 affiche une coopération quasi totale, lorsque l'on fait interagir 1 et 2 ensemble, cela réduit la contribution du joueur 1, à l'équilibre, de près de 15 % de sa valeur (de 9.22 à 8.12). Dans le même temps, le fait que le joueur 2 joue avec le joueur 1, plutôt qu'avec un partenaire qui lui est parfaitement semblable, permet de faire augmenter de près de 40 % son niveau de contribution (de 5.28 à 7.24), ce qui témoigne de l'importance du rôle du joueur 1, en tant que « leader » de l'effort coopératif. Au total, avec $a = \frac{3}{4}$, le niveau moyen de contribution d'un groupe composé uniquement de joueurs 1 passe de 9.22 à 7.68 lorsque l'on remplace un des individus par un joueur 2. Cependant, malgré l'effet globalement néfaste de l'introduction du joueur 2, sur le joueur le plus coopératif, en comparaison avec ce qui se passe dans le JBP standard, le mécanisme incitatif de récompense permet ici d'obtenir un niveau de coopération plus important entre ces deux joueurs, aux personnalités différentes et, en particulier, le joueur 2 est ici capable de satisfaire les attentes de son partenaire à travers sa décision de contribution.

Pour déterminer quels paramètres du modèle jouent, sans ambiguïté, sur le niveau de coopération, nous nous intéressons maintenant au calcul des dérivées.

Calcul des dérivées

En ce qui concerne les paramètres de capital de sympathie et de norme de contribution, étant donné que ceux-ci n'apparaissent pas dans l'expression de la fonction de récompense des agents, la dérivée de la contribution optimale, par rapport à ces paramètres, reste la même que dans le cas du JBP standard (cf. 5.2.1.3.2). En conséquence, on se limite à l'analyse du rôle de l'émotion de fierté.

La sensibilité à l'émotion de fierté ρ

$$\frac{\partial y_{r_i}^*}{\partial \rho_i} = \frac{\alpha_j}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} > 0 \text{ et } \frac{\partial y_{r_i}^*}{\partial \rho_j} = \frac{a\alpha_i^2}{2\beta_i(4\beta_i\beta_j - a^2\alpha_i\alpha_j)} > 0$$

Interprétation :

Comme dans le cas symétrique, le montant de contribution de chaque agent dépend positivement de l'intensité avec laquelle le sentiment de fierté est ressenti par l'un ou l'autre des joueurs, lorsqu'ils sont récompensés.

Par conséquent, notre modèle suggère que, toutes choses égales par ailleurs, l'augmentation de la sensibilité au sentiment de fierté, chez un seul agent, permet de promouvoir le niveau de coopération de l'ensemble des joueurs.

En ce qui concerne les autres paramètres, la complexité de l'interaction entre deux partenaires hétérogènes ne permet pas de réaliser une interprétation précise des dérivées calculées. Toutefois, nous pensons que la variation des niveaux de contribution, à l'équilibre, induite par une variation, à la marge, des paramètres de culpabilité et de sensibilité au comportement d'autrui, est à nouveau sensible à la manière dont les contributions, à l'équilibre initial, se comparent à la valeur des normes intériorisées par les deux agents.

De façon à mieux représenter ce qui se passe dans la réalité expérimentale, où les groupes de participants sont généralement composés de plus de deux joueurs, nous proposons maintenant une généralisation de notre modèle.

5.3. Généralisation

5.3.1. Le jeu du bien public dans sa version standard

Nous nous intéressons maintenant au jeu du bien public lorsqu'il est joué dans sa version *standard*, en un coup, par $n > 2$ joueurs. Le fait de généraliser notre formalisation nous permet d'analyser l'influence des émotions au niveau du groupe et d'obtenir des résultats qui peuvent être davantage comparables à ceux observés empiriquement, puisque la plupart des protocoles expérimentaux font participer plus de deux joueurs lors d'une même répétition du jeu.

On note toujours $I = \{1, 2, \dots, n\}$ l'ensemble des joueurs.

Le jeu se déroule conformément à ce qui est décrit plus haut (cf. 5.1.1.2.). On rappelle cependant, que pour tout $i \in I$, le gain matériel de l'agent i dans le jeu est donc défini comme :

$$\pi_i = d - y_i + a\left(\sum_{i \in N} y_i\right) \quad (5.3.1)$$

5.3.1.1. La fonction d'utilité

Pour tout agent $i \in I$, $i \neq j$, la fonction d'utilité est à nouveau définie comme suit :

$$U_i = AC_i + EA_{ij} \quad (5.3.2)$$

Nous supposons que l'utilité qui est liée à l'émergence des émotions auto-conscientes est toujours égale à $AC_i = \pi_i - \beta_i(\bar{y}_i - y_i)^2$, une fonction que nous avons déjà présentée précédemment (cf. 5.2.1.1.1). Ce choix nous semble relativement naturel étant donné que, s'agissant de normes de comportement *intériorisées*, on peut s'attendre à ce que l'introduction de nouveaux joueurs, conduisant à la formation d'un groupe, n'influe pas sur la manière dont l'individu est en mesure de déterminer la pertinence de ses propres actions (sous l'influence de l'émotion de culpabilité), sur le plan moral.

Formellement, l'originalité découlant de la généralisation de notre modèle, par rapport au JBP standard avec $n = 2$ joueurs, concerne la part de la fonction d'utilité qui est relative à l'évaluation d'autrui et aux émotions qui sont impliquées dans cette évaluation, puisque chaque agent $i \in I$ a désormais au moins deux partenaires.

La prise en compte d'autrui : l'insertion dans un groupe

Dans cette version généralisée du modèle, nous supposons que la fonction EA_{ij} est définie de la manière suivante :

$$EA_{ij} = \frac{1}{n-1} \sum_{j \neq i} \theta_{ij} \pi_j \quad (5.3.3)$$

Où θ_{ij} est toujours égal à $\theta_{ij} = \delta_i + \varepsilon_{ij}$, avec $\varepsilon_{ij} = \alpha_i(y_j - \bar{y}_i)$ (équations (5.2.3) et (5.2.5)).

Avant toute chose, il est nécessaire de souligner ici que, du fait de la définition du terme θ_{ij} , chaque joueur $i \in I$ est doté du même capital de sympathie vis-à-vis de chacun des autres agents qui forment, avec lui, le groupe dans lequel il joue. Ainsi, chaque joueur se sent aussi proche de chacun de ses partenaires (niveau intra individuel). Toutefois, il existe bien de l'hétérogénéité au niveau interindividuel, en ce qui concerne l'intensité avec laquelle un joueur se sent proche de ses partenaires. Cette précision ayant été apportée, nous passons maintenant à la présentation détaillée de la fonction EA .

La spécification de la fonction EA_{ij} nous permet d'introduire le rapport que chaque agent $i \in I$ entretient avec les individus qui composent le reste de son groupe. L'utilisation de la moyenne nous permet, d'une part, de prendre en considération le fait qu'un joueur tient compte des comportements individuels de chacun de ses partenaires au sein du groupe, lors de la formation de son appréciation de ceux-ci, mais aussi, d'autre part, de supposer que celui-ci est en mesure d'en tirer une information générale et synthétique, qui lui est utile lors de la prise de décision quant à son niveau de contribution. Cette approche nous permet ainsi de donner une certaine représentativité au groupe, en tant qu'entité au sein de laquelle l'individu s'exprime, lorsque celui-ci prend sa décision.

En outre, cette formalisation permet sans doute également de mieux appréhender la portée collective de l'acte de contribution, puisque dans le jeu, le fait de contribuer bénéficie à l'ensemble des membres du groupe. A l'inverse, le fait de supposer que l'individu tient compte de l'impact spécifique de sa contribution envers chacun des autres membres de son groupe, sans cet effort de synthèse, implique une certaine redondance qui semble contre-intuitive.

Pour mieux comprendre la manière dont nous conceptualisons ces éléments, nous procédons au développement de l'expression de la fonction EA_{ij} à partir de l'équation (5.2.3). On trouve alors, pour tout agent $i \in I$, $i \neq j$:

$$EA_{ij} = \frac{1}{n-1} \sum_{j \neq i} \delta_i \pi_j + \frac{1}{n-1} \sum_{j \neq i} \varepsilon_{ij} \pi_j \quad (5.3.4)$$

Le capital sympathie vis-à-vis du reste du groupe

Le premier terme de la fonction EA_{ij} telle qu'elle est définie dans l'équation (5.3.4) représente l'utilité que l'agent i dérive du fait de ressentir de la sympathie pour l'ensemble des autres membres de son groupe. Cette utilité est alors envisagée sous la forme d'une moyenne du plaisir que l'agent i retire de la perception qu'il a du bien-être (matériel) de chacun de ses partenaires, lequel est pondéré par la proximité qu'il ressent vis-à-vis de chacun d'eux.

Le recours à la moyenne permet d'introduire, chez l'agent, une certaine faculté à synthétiser, en tenant compte des issues du jeu pour ses différents partenaires, afin d'obtenir un indice général lui permettant de mesurer l'impact de son comportement de contribution sur la satisfaction qu'il retire du fait de promouvoir le bien-être des autres membres du groupe, et qui provient de la sympathie qu'il éprouve à leur égard.

De manière similaire à ce qui a été vu précédemment (cf. 5.2.1.1.2), ce sentiment de sympathie ressentie à l'égard du reste du groupe, par chacun des joueurs, se développe de manière inconditionnelle par rapport aux choix stratégiques mis en œuvre par ses partenaires et son influence, au moment de la prise de décision, s'effectue de concert avec l'influence de l'évaluation que chaque individu fait de la conduite morale des autres membres du groupe.

L'évaluation du comportement d'autrui dans le groupe

Le second terme de la fonction EA_{ij} désigne l'utilité que l'agent i retire de la valorisation qu'il fait de l'interaction qui se produit, pendant le jeu, entre lui et ses partenaires au sein du groupe, en partie sous l'effet de l'émergence des émotions d'évaluation d'autrui que sont la colère et la gratitude. Comme nous avons pu le voir dans le cas où il n'y a que deux joueurs, l'émergence de ces émotions chez un joueur provient de l'analyse de la façon dont le comportement de ses partenaires est en conformité avec la norme morale de contribution qu'il a intériorisée.

Les individus étant placés dans un groupe, chaque joueur analyse le comportement de l'ensemble de ses partenaires. Chacune de ces évaluations va procéder de concert avec l'émergence des émotions de colère (lorsqu'un partenaire n'a pas contribué suffisamment par rapport à la norme) et de gratitude (lorsqu'un (autre) partenaire a contribué davantage que ce que la norme préconise) qui vont participer à l'élaboration d'un (mé)contentement général de l'agent, vis-à-vis du reste de son groupe, en fonction de l'émotion qui est la plus élicitée (en fréquence et/ou en intensité) par le comportement des autres membres du groupe.

Lors de sa prise de décision, on suppose que chaque agent a besoin de synthétiser l'ensemble de ces différents ressentis émotionnels pour dégager un sentiment global, lui permettant d'évaluer son (mé)contentement à évoluer dans un tel groupe, ce qui représente ensuite une (dés)incitation à entreprendre un comportement coopératif.

Avec

$$U_i^C = AC_i + EA_{ij}$$

Où la fonction U_i^C représente alors l'utilité relative à la phase de contribution, qui a été exposée précédemment (cf. 5.3.1.1).

Ainsi, nous passons directement à la présentation de la fonction U_i^R :

$$U_i^R = R_{ij} + R_{ji} \quad (5.3.7)$$

Où, d'une part, R_{ij} se réfère à nouveau à l'utilité que l'agent i retire du fait de récompenser ses partenaires, avec :

$$R_{ij} = \sum_{j \neq i} (\theta_{ij} r_{ij} - \beta_i r_{ij}^2 - cr_{ij}) \quad (5.3.8)$$

Et où, d'autre part, R_{ji} se rapporte à l'utilité que l'agent i dérive du fait d'être récompensé par ses partenaires, avec :

$$R_{ji} = \sum_{j \neq i} r_{ji} + \frac{1}{n-1} \rho_i \sum_{j \neq i} r_{ji} \quad (5.3.9)$$

En somme, la fonction d'utilité de tout agent $i \in I$, $i \neq j$, peut également être écrite sous la forme suivante :

$$U_i = \pi_i - \beta_i (\bar{y}_i - y_i)^2 + \frac{1}{n-1} \sum_{j \neq i} \theta_{ij} \pi_j + \sum_{j \neq i} (\theta_{ij} r_{ij} - \beta_i r_{ij}^2 - cr_{ij}) + \left(1 + \frac{\rho_i}{n-1}\right) \sum_{j \neq i} r_{ji} \quad (5.3.10)$$

Comme dans les cas précédents, lorsque l'on pose $\alpha_i = \delta_i = \beta_i = \rho_i = 0$, on retrouve la représentation classique des agents économiques.

Afin d'apporter des précisions à la présentation de notre fonction d'utilité, nous nous intéressons premièrement à l'utilité liée à l'acte de récompenser (fonction R_{ij}), avant de nous concentrer, dans un deuxième moment, sur la description de l'utilité provenant du fait d'être récompensé (fonction R_{ji}).

5.3.2.1.1. L'utilité liée à l'acte de récompenser

La fonction R_{ij} (équation (5.3.8)) que nous présentons ici, n'est autre qu'une généralisation de celle qui avait été décrite dans le cas où $n = 2$ (cf. 5.2.2.1.1).

5.3.2.1.2. Une expression des sentiments réciproques à l'encontre de chaque partenaire

On retrouve, dans le premier terme de la somme qui constitue la fonction R_{ij} , la valorisation que le joueur i fait de l'acte de récompenser un de ses partenaires j . Ainsi, sur la base d'une analyse individualisée du comportement de tout autre agent j , le joueur i accorde un poids spécifique au fait de récompenser chaque partenaire en particulier, et ce phénomène se réalise pour l'ensemble des joueurs qui composent, avec lui, son groupe de jeu.

A nouveau, on remarque que l'agent i valorise d'autant plus positivement le fait de récompenser un partenaire j que celui-ci adopte un comportement qu'il évalue comme bienveillant, *i.e.* qui est en mesure d'éliciter l'émotion de gratitude de manière plus intense ($\theta_{ij} > 0$). Au contraire, lorsqu'un partenaire j ne répond pas aux attentes de l'agent i et contribue de sorte à le mettre en colère ($\theta_{ij} < 0$), celui-ci est conduit à se venger en s'abstenant de récompenser son partenaire.

5.3.2.1.3. Le coût lié à l'acte de récompenser son partenaire

Parallèlement, pour récompenser ses partenaires, l'agent i doit toujours supporter un coût c pour chaque unité offerte.

De plus, le coût psychologique lié à l'acte de récompenser l'autre joueur $-\beta_i r_{ij}^2$, est également toujours présent dans la fonction R_{ij} . Le recours à la présence d'une telle perte supplémentaire d'utilité, lorsqu'un agent récompense, est encore davantage justifié dans cette version du jeu.

En effet, nous supposons, à travers ce terme, que les agents ressentent une gêne, liée au sentiment de culpabilité, de plus en plus grande, à mesure qu'ils consacrent des ressources (trop) importantes afin de remercier un partenaire en particulier, alors même que ces ressources pourraient être nécessaires afin de se conformer à d'autres normes de comportement qu'ils ont intériorisées.

Dans cette déclinaison du jeu, il convient de remarquer que les ressources d'un agent ne sont pas infinies et que celui-ci a l'opportunité de récompenser l'ensemble de ses partenaires dans le jeu. Ainsi, afin de pouvoir se comporter raisonnablement, soit en étant en mesure de récompenser convenablement chacun de ses partenaires, mais aussi de conserver une partie de ses ressources pour lui-même, chaque individu contrebalance, en partie, son désir de réciprocité positive (lorsque celui-ci est suscité), avec une certaine aversion au fait de (trop) se démunir.

Parallèlement, ce coût psychologique peut également à nouveau s'interpréter comme une aversion des agents au fait de se placer en position de vulnérabilité lorsqu'ils récompensent un autre joueur, ou comme lié au risque de prévalence du sentiment de dette sur le sentiment de gratitude au moment du choix du niveau de récompense (cf. 5.2.2.1.1).

5.3.2.1.4. La satisfaction d'être reconnu, par le groupe, pour son comportement

Comparativement au cas où $n = 2$, la fonction R_{ji} (équation 5.3.9), mesurant l'utilité que l'agent i obtient lorsqu'il est récompensé par ses partenaires, se déduit systématiquement de celle qui avait été présentée plus haut (cf. 5.2.2.1.2), à l'exception d'une petite subtilité.

Le premier terme de cette fonction correspond à l'ensemble des gains matériels obtenus par l'agent i , du fait des différentes récompenses reçues de la part de ses partenaires, qui sont logiquement sommées.

La subtilité apparaît dans le deuxième terme de la fonction. Celui-ci est relatif au gain psychologique provenant du fait d'être récompensé ou, en d'autres mots, du fait de recevoir un certain niveau d'approbation de la part des autres membres du groupe, pour le comportement contributif que l'on a adopté.

Nous avons déjà discuté, plus haut, de la capacité des individus à synthétiser certaines informations, afin d'arriver à se former une idée (ou un sentiment) générale à l'égard du comportement de leurs partenaires dans le groupe. Ici, nous supposons que, lorsqu'un individu est récompensé par ses partenaires, un sentiment de fierté émerge en lui, sur la base des signes de valorisation qu'il reçoit, vis-à-vis de son comportement lors de la phase de contribution.

Cependant, nous supposons aussi que ce sentiment intervient après que l'agent ait synthétisé toute l'information qui lui est transmise. Pour ce faire, celui-ci tient compte uniquement du niveau d'approbation générale qui lui est communiqué par le reste du groupe, en se focalisant sur la moyenne des récompenses qu'il a reçues.

5.3.2.2. La résolution du jeu

Le jeu se résout à nouveau en recourant à un raisonnement par induction à rebours. Dans un premier temps, on détermine, pour chaque agent $i \in I$, une fonction de récompense qui définit le montant optimal à offrir à chacun de ses partenaires, en fonction du niveau de contribution de ceux-ci. Ensuite, en supposant que chaque agent utilise cette fonction de récompense, on peut déterminer, pour chaque agent, une fonction de meilleure réponse, en termes de comportement de contribution, qui permettent la détermination d'un unique équilibre de Nash sur les niveaux de contribution.

Ainsi, pour tout joueur $i \in I$, $i \neq j$, nous devons, en premier lieu, résoudre le programme de maximisation suivant (pour clarifier l'exposé, posons $i = 1$) :

$$\begin{array}{ll} \underset{r_{12}, r_{13}, \dots, r_{1n}}{\text{Max}} & U_1(r_{12}, r_{13}, \dots, r_{1n}) \\ \text{s.c.} & \\ & r_{1j} \geq 0 \quad \forall j \neq 1, j \in I \end{array}$$

5.3.2.3. Résultats

Nous procédons maintenant à l'analyse des résultats de notre modèle généralisé, en comparant les deux situations du JBP standard et du JBP avec phase de récompense.

Nous considérons alors un groupe de $n = 4$ joueurs. On suppose toujours que les dotations initiales des joueurs sont égales à $d = 10$ et que le coût à récompenser est de $c = 0.5$. Le tableau suivant résume les profils de paramètres des joueurs :

$i = \{1, 2, 3, 4\}$	δ_i	\bar{y}_i	β_i	α_i	ρ_i
Joueur 1	0.6	5	0.3	0.4	0.6
Joueur 2	0.4	4	0.5	0.5	0.5
Joueur 3	0.7	3	0.6	0.6	0.7
Joueur 4	0.5	4	0.4	0.5	0.4

TABLE 5.11.: Exemple de paramétrisation - Modèle généralisé

Il convient de noter que les joueurs 1 et 2 utilisés ici sont les mêmes que dans le cas à deux joueurs. Deux nouveaux profils différents (joueurs 3 et 4) de joueurs ont été arbitrairement créés et on suppose maintenant que ces quatre joueurs interagissent ensemble.

Avec de telles valeurs des paramètres, notre modèle prédit les résultats suivants :

JBP standard avec $a = 0.5$				
y_1^*	y_2^*	y_3^*	y_4^*	Moyenne du groupe
4.13	3.60	3.07	3.56	3.59
$y_1^* < \bar{y}_1$	$y_2^* < \bar{y}_2$	$y_3^* < \bar{y}_3$	$y_4^* < \bar{y}_4$	

TABLE 5.12.: Résultats - JBP standard - Modèle généralisé

JBP avec récompense $a = 0.5$ et $c = 0.5$				
$y_{r_1}^*$	$y_{r_2}^*$	$y_{r_3}^*$	$y_{r_4}^*$	Moyenne du groupe
8.39	6.55	5.79	6.96	6.92
$y_{r_1}^* > \bar{y}_1$	$y_{r_2}^* > \bar{y}_2$	$y_{r_3}^* > \bar{y}_3$	$y_{r_4}^* > \bar{y}_4$	

		j				Total Versé
		r_{ij}^*	1	2	3	
i	1	—	1.20	0.69	1.47	3.36
	2	2.09	—	0.79	1.38	4.26
	3	2.86	1.94	—	2.15	6.95
	4	2.75	1.60	1.12	—	5.47
Total reçu		7.7	4.74	2.6	5	20.04

TABLE 5.13.: Résultats - JBP récompense - Modèle généralisé

Avec cette paramétrisation, on constate à nouveau que l'opportunité de récompense est utilisée par chacun des joueurs et qu'elle a un effet bénéfique du point de vue du niveau de coopération, lors de la phase de contribution. En effet, le niveau moyen de contribution augmente sensiblement et passe de 3.59, dans le jeu standard, à 6.92 dans le JBP avec récompense.

Un autre résultat intéressant de notre modèle est ici que le joueur contribuant le moins à l'équilibre (3) est également celui qui est le moins récompensé tandis que le joueur qui contribue le plus (1) est celui qui obtient le plus grand montant de récompense. Ces deux résultats sont cohérents avec ce qui est observé dans l'expérience de Rand *et al.* (2009). Toutefois, notre modèle prédit, dans cette situation, que les niveaux de récompense les plus forts proviennent majoritairement de l'agent qui contribue le moins (3), ce qui peut sembler contre-intuitif. La raison expliquant un tel résultat est que, comparativement à la norme de contribution du joueur 3 (qui est de 3), les contributions relativement importantes des autres joueurs déclenchent en lui un niveau important de gratitude, ce qui conduit donc à d'importants niveaux de récompense de sa part.

5.3.2.3.1. Introduction d'une possibilité de communication

Imaginons maintenant, que l'on autorise les agents à communiquer entre eux avant le début du jeu. Sur la base des éléments de revue que nous avons mentionnés, en ce qui concerne les résultats empiriques dans le JBP, il semble que le fait d'autoriser une telle communication permette aux individus de se coordonner sur un comportement plus coopératif. Nous supposons que, dans notre modèle, une façon de représenter l'influence de la communication est de supposer que celle-ci permet à tous les agents de prendre leur décision en utilisant la même norme de contribution. Nous supposons également que le joueur pour qui la norme de contribution est la plus importante, est à l'initiative, durant la phase de communication, afin de convaincre ses partenaires des bienfaits de la coopération. Cependant, on suppose tout d'abord que la communication ne s'effectue pas en face à face. Dans ces conditions, on peut supposer que la force de persuasion du joueur 1 est limitée. Par conséquent, dans notre exemple, on suppose que le joueur 1 continue à utiliser sa norme de contribution qui vaut 5, tandis que ses partenaires se coordonnent sur

la valeur moyenne des normes dans notre échantillon, qui est 4 (les autres valeurs des paramètres restant les mêmes). Lorsque la communication avant le début du jeu est possible, on obtient les résultats suivants (les moyennes de groupe obtenues sans communication sont indiquées entre parenthèse) :

JBP standard - Communication - $a = 0.5$ et $\bar{y} = 5$				
y_1^*	y_2^*	y_3^*	y_4^*	Moyenne du groupe
4.24	3.68	3.84	3.66	3.86 (3.59)
$y_1^* < \bar{y}_1$	$y_2^* < \bar{y}_2$	$y_3^* < \bar{y}_3$	$y_4^* < \bar{y}_4$	

TABLE 5.14.: Résultats - JBP standard - Communication

JBP récompense - Communication - $a = 0.5$ et $\bar{y} = 5$				
$y_{r_1}^*$	$y_{r_2}^*$	$y_{r_3}^*$	$y_{r_4}^*$	Moyenne du groupe
8.50	6.63	6.57	7.05	7.18 (6.92)
$y_{r_1}^* > \bar{y}_1$	$y_{r_2}^* > \bar{y}_2$	$y_{r_3}^* > \bar{y}_3$	$y_{r_4}^* > \bar{y}_4$	

TABLE 5.15.: Résultats - JBP récompense - Communication

Par conséquent, nos résultats indiquent, dans cet exemple, que la possibilité de communication entre les joueurs, ayant pour conséquence une coordination de chaque membre du groupe sur une même valeur de la norme de contribution, permet d'augmenter le niveau moyen de coopération à l'équilibre, notamment dans le cas du JBP avec phase de récompense.

Expérimentalement, la forme de communication qui permet le plus de promouvoir la coopération est le face à face. Nous supposons que l'avantage d'une discussion en face à face, sur les autres modes de communication possible, est qu'elle est probablement capable de susciter un sentiment plus important de proximité entre les individus, tout en leur permettant de se coordonner sur une norme de comportement plus coopérative.

Ainsi, imaginons maintenant que nos quatre joueurs ont la possibilité de communiquer entre eux, en face à face, avant le début du jeu. Dans ces conditions, nous supposons que cela entraîne une augmentation du paramètre δ (disons) de 0.1, pour chacun des joueurs, du fait que ceux-ci sont amenés à se sentir plus proches les uns des autres. De fait, les nouvelles valeurs du paramètre δ , pour chaque joueur, sont les suivantes :

$i = \{1, 2, 3, 4\}$	δ_i
Joueur 1	0.7
Joueur 2	0.5
Joueur 3	0.8
Joueur 4	0.6

5. Le Cas Statique

Parallèlement, on suppose que la possibilité de communication entraîne une coordination de chacun des joueurs sur la valeur de la norme de contributions la plus élevée au sein du groupe. Par conséquent, on suppose pour chacun des joueurs $\bar{y} = 5$. Dans ces circonstances, notre modèle prédit les résultats suivants (les moyennes de groupe obtenues sans communication sont indiquées entre parenthèse) :

JBP standard - Face à face - $a = 0.5$ et $\bar{y} = 5$				
y_1^*	y_2^*	y_3^*	y_4^*	Moyenne du groupe
4.66	4.68	4.83	4.66	4.70 (3.59)
$y_1^* < \bar{y}_1$	$y_2^* < \bar{y}_2$	$y_3^* < \bar{y}_3$	$y_4^* < \bar{y}_4$	

TABLE 5.16.: Résultats - JBP standard - Face à face

JBP récompense - Face à face - $a = 0.5$ et $\bar{y} = 5$				
$y_{r_1}^*$	$y_{r_2}^*$	$y_{r_3}^*$	$y_{r_4}^*$	Moyenne du groupe
8.92	7.63	7.56	8.06	8.04 (6.92)
$y_{r_1}^* > \bar{y}_1$	$y_{r_2}^* > \bar{y}_2$	$y_{r_3}^* > \bar{y}_3$	$y_{r_4}^* > \bar{y}_4$	

TABLE 5.17.: Résultats - JBP récompense - Face à face

Conformément avec ce que l'on retrouve expérimentalement, notre exemple illustratif suggère que la communication en face à face, qui permet le mieux aux agents de se coordonner sur un comportement coopératif, ainsi que de faire augmenter le sentiment de proximité entre les individus, est le mode de communication le plus efficace dans la promotion de la coopération entre des participants au JBP, notamment lorsque l'opportunité de récompense est offerte aux agents.

5.3.2.3.2. Simulations : comportement moyen de groupes générés aléatoirement

Etant donné le nombre important de paramètres dans notre modèle, il est compliqué d'obtenir des informations générales à partir d'une ou plusieurs paramétrisations particulières. Ainsi, pour nous assurer que nos résultats restent cohérents, pour d'autres valeurs des paramètres que celles que nous avons utilisées en exemple, nous nous proposons d'observer la contribution moyenne que l'on obtient, pour des groupes composés de $n = 4$ agents dont les paramètres ont été tirés dans des distributions aléatoires (normales). Pour chaque groupe aléatoirement généré, on calcule, on détermine l'équilibre de Nash du jeu, ce qui nous permet ensuite d'obtenir un résultat moyen, sur le nombre voulu de groupes.

Tirages aléatoires des paramètres

Pour chaque agent aléatoirement généré, on tire l'ensemble de ses paramètres dans des distributions normales qui peuvent avoir différentes moyennes ou qui peuvent être ou non asymétriques. Pour commencer, intéressons nous aux deux paramètres β et ρ . On rappelle que, pour tout $i \in I$, on suppose $\beta_i \in]0, 1[$ et $\rho_i \in [0, 1]$. En conséquence, ces deux paramètres sont tirés dans une distribution normale de moyenne $\mu = 0.5$ et d'écart type $\sigma = 0.125$, dont la fonction de densité de probabilité est représenté dans le graphique suivant :

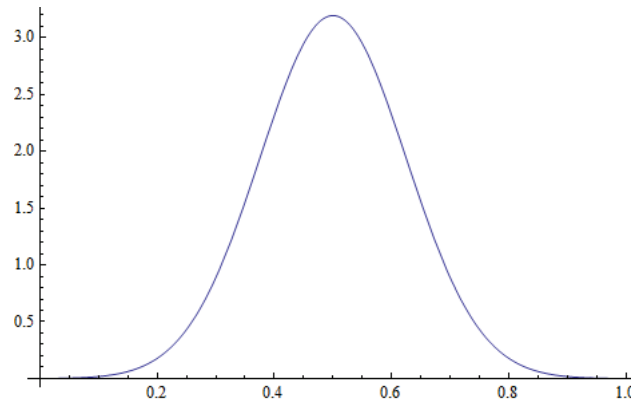


FIGURE 5.3.1.: Fonction de densité de probabilité - Distribution normale $\mu = 0.5$ et $\sigma = 0.125$

Comme on peut le voir dans le graphique 5.3.1, l'intérêt de cette distribution est qu'elle couvre l'intégralité des valeurs possibles pour les deux paramètres en question. En outre, du fait des caractéristiques des distributions de loi normale, on sait que les paramètres tirés ont une grande probabilité d'être proches des valeurs centrales de l'ensemble de définition $]0, 1[$ et une faible probabilité de se retrouver dans des valeurs extrêmes, ce qui nous semble cohérent par rapport à la réalité.

En ce qui concerne le paramètre α , on suppose, pour tout $i \in I$, que $\alpha_i \in]\beta_i, 2\beta_i[\cap]0, 1[$. En conséquence, pour nos simulations, nous avons choisi de tirer, pour tout agent $i \in I$, le paramètre α_i dans une distribution normale asymétrique de moyenne $\mu_\alpha = \frac{\beta_i + \min(2\beta_i, 1)}{2}$ (ce qui correspond au milieu de son ensemble de définition), d'écart-type $\sigma = 0.125$ et dont le paramètre d'asymétrie vaut -3 . Nous introduisons ici une asymétrie, de sorte à éviter que le paramètre α ne prenne trop souvent des valeurs extrêmes en comparaison de la valeur de β .

Par ailleurs, on suppose pour tout $i \in I$ que $\delta_i \in [0, 1]$. Il pourrait ainsi sembler logique de tirer aléatoirement les valeurs de δ dans la même distribution que celle dans laquelle on tire les paramètres β et ρ . Toutefois, nous considérons que ce paramètre est représentatif de « l'orientation sociale » des individus (cf. 5.2.1.1.2) et il ressort des études empiriques qu'il y a une faible représentativité des individus altruistes dans la population, comparativement à la proportion d'individualistes et de coopérateurs que l'on y trouve. Par conséquent, pour tout agent $i \in I$, on tire les valeurs de δ_i dans une distribution normale asymétrique de moyenne $\mu = 0.5$, d'écart type $\sigma = 0.125$ et dont le paramètre d'asymétrie

5. Le Cas Statique

est égal à -10 . La fonction de densité de probabilité de cette distribution est représentée dans le graphique suivant :

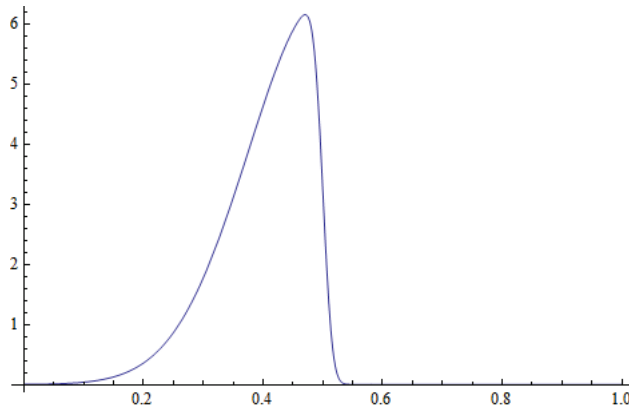


FIGURE 5.3.2.: Fonction de densité de probabilité - Distribution normale asymétrique $\mu = 0.5$, $\sigma = 0.125$ et $\alpha = -10$

Enfin, en posant, pour tout agent $i \in I$, $\bar{y}_i = P_i \times d$, il convient également de tirer aléatoirement le paramètre $P_i \in [0, 1]$. Dans le cadre de nos simulations, nous choisissons de procéder au tirage de ce paramètre dans une distribution normale asymétrique de moyenne $\mu = 0.5$, d'écart type $\sigma = 0.125$ et dont le paramètre d'asymétrie est égale à -2 . Une légère asymétrie est introduite ici, de sorte à prendre en compte le fait que des valeurs extrêmement grandes de la norme de contribution des individus nous semblent plus rares que des valeurs extrêmement faibles, en raison de la présence systématique d'individus adoptant des comportements de « passager clandestin » dans les données empiriques. La fonction de densité de probabilité de cette distribution est représentée dans le graphique suivant :

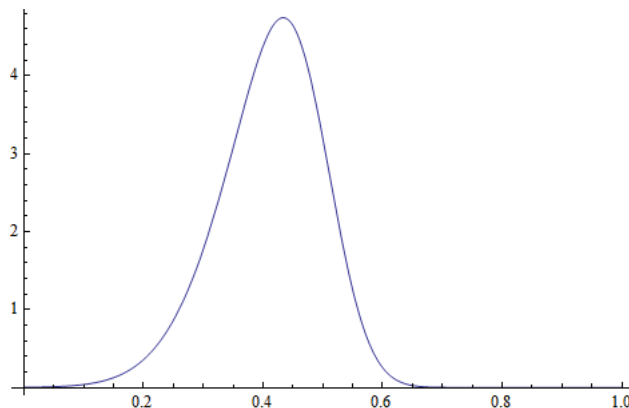


FIGURE 5.3.3.: Fonction de densité de probabilité - Distribution normale asymétrique $\mu = 0.5$, $\sigma = 0.125$ et $\alpha = -2$

Au total, pour chaque agent généré dans nos simulations, on tire aléatoirement chacun de ses paramètres dans les distributions que l'on vient de décrire et on les fait jouer au JBP standard ou avec récompense par groupe de $n = 4$.

Résultats

Pour nos simulations, nous nous intéressons à la fois au JBP standard et au JBP avec phase de récompense, joués avec deux valeurs différentes du rendement marginal du bien public. En conséquence, nous avons généré 100 groupes de $n = 4$ joueurs pour participer à chacune de nos quatre « conditions ». Le tableau suivant résume les résultats que nous avons obtenus :

JBP standard		
	$a = 0.5$	$a = 0.75$
\bar{y}^*	3.64	4.22

JBP récompense		
	$a = 0.5$	$a = 0.75$
\bar{y}_r^*	7.03	8.67
\bar{r}^*	1.62	2.69

TABLE 5.18.: Modèle généralisé - Simulations sur 100 groupes aléatoirement générés

Dans le tableau 5.18, les notations \bar{y}^* et \bar{y}_r^* désignent la moyenne des contributions optimales, calculée sur l'ensemble des joueurs de chaque condition (soit $4 \times 100 = 400$ agents), que l'on obtient respectivement dans le JBP standard et dans le JBP avec récompense. Par ailleurs, on note \bar{r}^* la moyenne, calculée sur l'ensemble des joueurs, du montant total utilisé par un joueur pour récompenser ses $n - 1$ partenaires.

Ces résultats sont à l'image de ceux obtenus jusqu'ici, ce qui tend à démontrer la pertinence des exemples illustratifs que nous avons proposés précédemment. Ainsi, on constate qu'en moyenne, les contributions optimales sont plus importantes dans le JBP avec récompense que dans le JBP standard, du fait que les individus usent de l'opportunité de récompense pour communiquer leur gratitude et leur approbation de manière ciblée. En outre, on voit également que le niveau de coopération est stimulé, dans le JBP standard et plus encore dans le JBP avec récompense, par une augmentation du rendement marginal du bien public, ce qui indique une certaine efficacité du mécanisme incitatif de subvention des contributions individuelles, lorsque le jeu est joué en un coup.

5.4. Conclusion

Au total, l'analyse des résultats obtenus dans ce chapitre montrent comment l'intégration de l'influence des émotions morales dans un modèle économique, qui reste dans le cadre de la théorie des jeux et du choix rationnel, permet d'obtenir des prédictions qui sont plus en accord avec ce qui est observé expérimentalement, mais également de poursuivre la recherche des facteurs déterminants de la coopération.

En effet, les résultats de notre modèle suggèrent que le mécanisme incitatif de récompense est efficace pour promouvoir la coopération des agents dans le jeu du bien public joué en un coup, ce qui est cohérent avec les données expérimentales que nous avons précédemment examinées. De plus, la justification théorique offerte à ces résultats, par notre modèle, par sa conception même, est en accord avec les éléments d'explications des « patterns » expérimentaux que nous avons revus dans le chapitre 4 (cf. 4.2.3).

Ainsi, dans la version standard du jeu du bien public, le manque d'incitation économique à coopérer tend à entraîner l'émergence d'émotions négatives lorsque les agents évaluent le comportement de leurs partenaires. En particulier, lorsqu'un agent qui a de fortes dispositions à coopérer est amené à jouer avec un individu plus égoïste, celui-ci est contrarié par le manque d'effort de son partenaire et, sous l'effet d'émotions comme la colère ou le dégoût, sa motivation à contribuer est amoindrie, notamment à travers la stimulation d'un désir de réciprocité négative. Face à cela, accroître l'attrait économique du bien public, par l'augmentation de son rendement marginal (ou par l'octroi de subventions aux contributions individuelles) se révèle être une solution permettant de limiter la portée du problème lié à la présence d'agents ayant tendance à adopter des comportements de « passager clandestin ». Nos simulations sur des groupes générés aléatoirement montrent ainsi que le niveau moyen de contribution augmente de près de 16 % (de 3.64 à 4.22) lorsque le rendement marginal du bien public passe de $a = 0.5$ à $a = 0.75$.

Parallèlement, lorsque l'option de récompense est disponible, cela offre aux agents la possibilité d'avoir une interaction ciblée avec chacun de leurs partenaires. Cette interaction ciblée sert alors de canal aux individus afin de communiquer leur gratitude, ainsi que leur valorisation (reconnaissance) des efforts consentis, aux membres de leur groupe qui le méritent, sous la forme d'une récompense monétaire (dans notre version du jeu). Ainsi, cela offre une incitation économique supplémentaire à contribuer, pour les agents qui cherchent à être récompensés pour des considérations d'ordre matériel. Par ailleurs, cela représente aussi une forte incitation pour les agents qui cherchent à obtenir l'approbation de leurs partenaires, pour déclencher l'émergence en eux du sentiment de fierté. Par conséquent, l'ensemble de ces motivations supplémentaires à contribuer, ainsi que la pluralité des profils d'individus concernés par celles-ci, permet de susciter une plus grande coopération des agents lorsqu'une phase de récompense est ajoutée au déroulement du jeu du bien public. En témoigne notamment le fait que, dans nos simulations sur des groupes aléatoirement générés, le niveau de contribution est presque doublé (3.64 à 7.03) lorsque la possibilité de récompense est introduite, comparativement à ce qui est obtenu dans la version standard du jeu (le rendement marginal du bien public étant dans chaque cas égal à $a = 0.5$).

L'ensemble des résultats que nous avons dégagés dans ce chapitre sont représentatifs de ce qui se passe dans le cas d'un jeu du bien public joué en un coup. Ainsi, nos différentes analyses nous ont permis d'apprécier les effets qui sont liés à l'augmentation du rendement marginal ou à l'introduction d'une phase de récompense, sur le niveau de coopération des agents. Ces résultats sont intéressants en tant que tels, mais ils le sont cependant d'autant plus si on peut les étendre à une perspective de long terme, en déterminant comment ces différents mécanismes incitatifs sont à même d'influer sur la dynamique de coopération. En effet, on peut supposer que le comportement effectif des agents, lors de la première période, peut avoir une dimension déterminante sur la dynamique qui s'instaure par la suite au sein du groupe, sous l'influence des différentes émotions qui sont suscitées chez les agents, en fonction de la capacité de leurs partenaires à satisfaire leurs attentes.

Ainsi, lors du prochain chapitre, nous allons analyser comment les différentes versions du jeu du bien public sont susceptibles de provoquer des dynamiques différentes de comportements de contribution et de relations interindividuelles, lorsque les agents jouent de manière répétée, au sein d'un même groupe.

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

Dans ce chapitre, nous nous intéressons à la dynamique des relations sociales et, plus précisément, à la manière dont les interactions qui surviennent au sein d'un groupe d'individus sont capables de façonner l'évolution de ces relations. Une des clés de la réussite d'un groupe devant mettre en œuvre un effort coopératif, se trouve dans la capacité des individus qui le composent, à se rapprocher socialement les uns des autres, jusqu'à ce que chaque membre intègre l'objectif commun au groupe comme un objectif personnel significatif et se soucie véritablement du bien-être de ses partenaires, en le considérant comme un prolongement important de son propre bien-être.

Nous pensons que les individus ont par nature des dispositions (qui peuvent probablement être modulées par la volonté, l'humeur du moment, le contexte etc.) leur permettant de se sentir (plus ou moins) proches d'agents inconnus, lorsqu'on les insère dans un nouveau groupe. De plus, nous considérons également que ceux-ci ont développé des stratégies de « référence », en s'appuyant sur leur perception des normes de comportement qu'ils jugent importantes. Ces deux éléments expliquent sans doute pourquoi on observe des comportements coopératifs entre des êtres humains qui viennent de se rencontrer. Par ailleurs, on considère que le sentiment de proximité et de sympathie, qui se caractérise différemment chez chaque agent d'un même groupe, au départ de l'interaction, évolue individuellement au cours du temps, au gré des choix et des actions qui sont entreprises par les partenaires d'un même groupe.

De manière similaire à la conception proposée par Van Dijk et Van Winden (1997) (cf. 3.2.3), nous considérons que, lors d'interactions répétées entre les agents d'un même groupe, ce qui est déterminant dans l'évolution des relations entre ces individus, ce sont les émotions qui sont élicitées au cours de chacune des répétitions de l'interaction. En effet, les émotions sont marquantes et participent aux jugements que les agents produisent à propos des autres et de leurs comportements. Elles sont de puissants indicateurs de ce qui anime véritablement les agents et elles les guident lors de leurs choix, y compris économiques.

Les émotions ont donc tout naturellement un impact dans l'évolution dynamique des relations, en ce sens qu'elles prodiguent une certaine capacité de mémorisation d'un événement marquant pour un agent, qui s'étend également aux individus particuliers qui sont impliqués dans un tel événement. Les émotions peuvent aussi participer à l'élaboration d'une certaine catégorisation d'un individu, ou plus largement d'un groupe d'individus, après un événement, ou une série d'événements, marquants, liée à ces personnes.

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

Par exemple, la colère et la déception ressentie à l'encontre d'un collaborateur qui n'a pas répondu à nos exigences peut nous conduire à le placer dans la catégorie des personnes à éviter et peut ensuite entraver la bonne poursuite (voire la possibilité) de futures interactions avec lui, du fait de considérations de vengeance, ou encore d'une perte de motivation ou d'intérêt pour le fait d'interagir avec quelqu'un qui nous a déçu. Ce phénomène peut potentiellement continuer indéfiniment, du moins jusqu'à ce que la personne en question donne la preuve qu'elle est à nouveau digne de confiance lors d'une ou plusieurs nouvelles interactions.

A l'inverse, l'émergence du sentiment de gratitude suite à l'adoption d'un comportement exemplaire par un partenaire, durant l'élaboration d'un projet commun, nous aide à catégoriser cette personne comme quelqu'un de fiable, avec qui on a envie d'interagir à nouveau et vis-à-vis de qui on peut développer un fort désir d'engagement conduisant à s'investir d'autant plus lorsque l'occasion de travailler à nouveau avec lui se présente. Là encore, nous pensons que pour qu'un tel phénomène dynamique s'enraye, cela nécessiterait que l'individu en question se montre décevant dans son comportement lors d'une ou plusieurs nouvelles interactions.

Ainsi, notre étude du rôle des émotions morales dans la prise de décision humaine, qui a mis en lumière l'influence spécifique des émotions morales, nous permet d'envisager et de cibler quelles sont celles qui sont en mesure d'influencer la dynamique des relations interpersonnelles.

Sur la base de ces réflexions, les prochaines sections présentent une approche formelle de l'évolution dynamique de la relation entre les agents dans le jeu du bien public, qui s'inscrit directement dans le prolongement du modèle original présenté dans le chapitre 5. Ce travail se place également dans la continuité de l'approche de Van Dijk et Van Winden (1997), laquelle met le rôle des émotions au centre de la dynamique de la formation et de l'évolution des liens sociaux. Néanmoins, dans leur article, ceux-ci regrettaient l'absence d'une théorie fondamentale des relations interpersonnelles qui prendrait véritablement en compte la présence de composantes affectives et émotionnelles. En ce sens, notre contribution est de pousser plus loin l'exploration du rôle des émotions, en s'intéressant de manière ciblée aux sentiments spécifiques qui sont susceptibles d'avoir une influence dynamique sur l'évolution des relations interpersonnelles, dans les situations de dilemme social.

6.1. Formalisation

L'objet de ce travail formel, est d'analyser le comportement économique des agents dans les différentes déclinaisons du jeu du bien public que nous avons décrites au cours du chapitre 5 (cf. 5.1), lorsqu'ils jouent de manière répétée, un nombre fini de fois et avec les mêmes partenaires (le groupe étudié reste composé des mêmes agents à chaque répétition du jeu). En conséquence, nous utilisons les mêmes notations et les mêmes concepts que ceux qui ont été présentés dans le chapitre précédent.

Etant donné la multiplicité des interactions à considérer, dans l'ensemble des situations que nous analysons, ajouter un degré supplémentaire de difficulté, en s'intéressant à l'évolution dynamique de la relation entre les joueurs, est un défi qui nous a conduit à adopter une méthodologie tournée principalement vers la simulation. Néanmoins, dans un premier temps, nous présentons les hypothèses théoriques relatives au fonctionnement dynamique de notre modèle.

6.1.1. Hypothèses

Nous supposons, dans notre modèle, que lorsqu'un groupe d'agents joue de manière répétée, T fois, au jeu du bien public, la relation entre ces agents est affectée par les interactions qui se déroulent à chaque période, sous l'influence des émotions que ces interactions suscitent. Plus précisément, nous supposons que l'émergence des émotions d'évaluation d'autrui, que sont d'une part la colère ou la déception, et d'autre part la gratitude, affectent le capital de sympathie dont chaque individu dispose, à l'encontre de ses partenaires, et ainsi que ces émotions sont en mesure de modifier le niveau de proximité entre les individus d'un même groupe.

Il convient de noter que tous les éléments formels qui sont exposés dans cette section s'appliquent, sans distinction, au jeu du bien public standard comme au jeu du bien public avec phase de récompense.

6.1.1.1. Le cas du couple

Supposons, dans un premier temps, qu'un couple d'agents participe au jeu du bien public, de manière répétée.

Pour tout agent $i \in I$, $i \neq j$, le terme qui est au cœur de la dynamique d'évolution des relations interpersonnelles est :

$$\theta_{ij} = \delta_i + \varepsilon_{ij} \quad (6.1.1)$$

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

Ce terme se compose, en effet, des deux éléments centraux à notre conceptualisation de l'évolution dynamique des relations. Comme cela est détaillé dans le chapitre 5, δ_i représente le capital de sympathie dont l'agent i dispose envers tout partenaire avec lequel il est susceptible d'interagir, tandis que $\varepsilon_{ij} = \alpha_i(y_j - \bar{y}_i)$ est un indice du jugement moral que porte i sur le comportement adopté par son partenaire j , sous l'influence des émotions d'évaluation d'autrui.

Lors de la période initiale, $t = 1$, on pose, pour chaque agent $i \in I$, $\delta_i = \delta_{i,1}$, où $\delta_{i,1}$ est une constante exogène qui correspond au capital initial de sympathie dont dispose l'agent i vis-à-vis de son partenaire dans le jeu.

A chaque période t , lors d'une répétition du jeu, chaque partenaire détermine son niveau de contribution optimal y^* au bien public. Sur cette base, chaque agent i est en mesure d'évaluer la convenance du comportement contributif de son partenaire j , à l'aide de la fonction suivante :

$$\varepsilon_{ij,t}^* = \alpha_i(y_{j,t}^* - \bar{y}_i) \quad (6.1.2)$$

D'une période à la suivante, on suppose alors que cette évaluation fait évoluer le capital de sympathie initial, selon la relation suivante :

$$\delta_{i,t+1} = \delta_{i,t} + f(\varepsilon_{ij,t}^*, \delta_{i,t}) \quad (6.1.3)$$

Avec

$$f(\varepsilon_{ij,t}^*, \delta_{i,t}) = \frac{e^{\varepsilon_{ij,t}^*} - \frac{1-\delta_{i,t}}{1+\delta_{i,t}}}{e^{\varepsilon_{ij,t}^*} + \frac{1-\delta_{i,t}}{1+\delta_{i,t}}} - \delta_{i,t} \quad (6.1.4)$$

Cette fonction f , qui est semblable à celle qui est fournie en annexe dans le travail de van Dijk et van Winden (1997), possède les propriétés suivantes :

1. $f(0, \delta_{i,t}) = 0$
2. $\frac{\partial f(\varepsilon_{ij,t}^*, \delta_{i,t})}{\partial \varepsilon_{ij,t}^*} > 0$
3. $\lim_{\varepsilon^* \rightarrow +\infty} f(\varepsilon^*, \delta) = 1 - \delta$
4. $\lim_{\varepsilon^* \rightarrow -\infty} f(\varepsilon^*, \delta) = -1 - \delta$

Les deux premières propriétés de la fonction f décrivent la manière dont nous conceptualisons l'évolution dynamique des relations interpersonnelles dans notre modèle, tandis que les deux dernières permettent de délimiter l'espace au sein duquel cette évolution s'opère.

La propriété 1 signifie que si l'interaction à la période t , laisse l'agent i indifférent ou, tout du moins, dans un état émotionnel neutre, soit $\varepsilon_{ij,t}^* = 0$, alors le capital de sympathie du joueur i reste identique entre t et $t + 1$. Ainsi, si la contribution du joueur j se conforme à la norme du joueur i ($\varepsilon_{ij,t}^* = 0 \Leftrightarrow y_j^* = \bar{y}_i$), celui-ci éprouvera toujours la même sympathie à son égard à la période suivante.

La propriété 2 permet de comprendre ce qu'il se passe à partir du moment où l'agent j ne répond pas exactement aux exigences de l'agent i lors de la période t . En effet, cette propriété implique que f est une fonction croissante de l'évaluation $\varepsilon_{ij,t}^*$ réalisée par l'agent i .

Des deux premières propriétés, on déduit que pour tout $\varepsilon_{ij,t}^* > 0$ ($\Leftrightarrow y_j^* > \bar{y}_i$), on a $f(\varepsilon_{ij,t}^*, \delta_{i,t}) > 0$, ce qui implique, d'après l'équation (6.1.3), $\delta_{i,t+1} > \delta_{i,t}$.

Parallèlement, pour tout $\varepsilon_{ij,t}^* < 0$ ($\Leftrightarrow y_j^* < \bar{y}_i$), on a $f(\varepsilon_{ij,t}^*, \delta_{i,t}) < 0$, ce qui conduit à $\delta_{i,t+1} < \delta_{i,t}$.

Avant de nous intéresser aux propriétés 3 et 4 de la fonction f , nous interprétons en détails les implications relatives aux deux premières propriétés.

6.1.1.1.1. Les bienfaits dynamiques de la gratitude

Lors de la période t , si l'agent j contribue davantage que ce qui est attendu par i , la gratitude ressentie par ce dernier ($\varepsilon_{ij,t}^* > 0$) conduit à l'augmentation de δ_i , d'une période à l'autre, soit à un accroissement du « capital affectif » dont dispose j , auprès de i .

Ce capital, à travers sa modification dans le temps, peut alors être envisagé comme un indice utile à la mémorisation, puisqu'il permet à l'agent i de tenir compte, en $t + 1$, du fait qu'il a positivement évalué l'interaction qui a eu lieu entre lui et son partenaire, lors de la période t .

En conséquence de cette augmentation du capital δ_i , lors de la répétition suivante du jeu (en $t + 1$), l'agent i se sent plus proche de l'agent j , ce qui donne lieu à plusieurs implications.

La première est que l'agent i accorde un poids inconditionnel d'autant plus grand au sort de son partenaire dans le jeu, ce qui contribue à faciliter l'adoption, par i , d'un comportement prosocial, lors de cette nouvelle période.

En ce sens, l'augmentation de δ_i oriente l'agent i vers l'adoption d'un comportement agissant en faveur d'une bonne entente (présente, et potentiellement future, en vertu de la dynamique positive qu'un comportement prosocial peut induire) avec son partenaire, plutôt que vers un comportement égoïste lui permettant de profiter immédiatement de plus de ressources matérielles. Comme nous le verrons ci-après, les agents, dans notre

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

modèle, ne réalisent pas de maximisation intertemporelle. On ne pose pas non plus d'hypothèse spécifique sur leur préférence pour le présent. Néanmoins, le raisonnement que nous venons de décrire est illustratif de la capacité de notre modèle à prendre en compte le fait que l'émotion de gratitude est susceptible de renforcer le niveau de patience économique des agents (cf. 2.2.3.1), en les guidant vers l'adoption de comportements qui, bien qu'étant moins attractifs sur le plan matériel, représentent un intérêt dans la formation d'une relation pérenne avec ses partenaires, qui peut lui être bénéfique dans le futur.

De plus, l'augmentation de δ_i implique également une résistance accrue de l'agent i face aux conséquences émotionnelles négatives liées à l'adoption d'un comportement inapproprié, de la part de j , en $t + 1$. Par ce dernier élément, nous mettons en lumière comment notre modèle intègre l'influence dynamique de la gratitude dans l'évolution de la capacité des individus à faire confiance, à être indulgents et à pouvoir pardonner (cf. 2.2.3.2).

Enfin, de manière imagée, on peut dire que l'accumulation d'un « capital affectif positif », par l'agent j , auprès de l'agent i , permise par l'évolution dynamique de δ_i , représente un transfert intertemporel des sentiments positifs ressenti (en t) par l'agent i , à l'encontre de son partenaire.

6.1.1.1.2. La spirale négative de la colère

A l'inverse, si lors de la période t , l'agent i juge que le comportement de son partenaire n'est pas approprié, celui-ci se retrouve alors sous l'influence de la colère et/ou de la déception ($\varepsilon_{ij,t}^* < 0$) et ces sentiments provoquent une dégradation de la sympathie qu'il porte à son partenaire. Cette fois, le « capital affectif » dont dispose j , auprès de i , s'amoindrit de t à $t + 1$ et la nouvelle valeur de δ_i garantit que l'agent i garde en mémoire, lors de la nouvelle période, le fait qu'il a évalué négativement son interaction précédente avec j .

De fait, ceci implique l'installation, chez i , d'une certaine distance entre lui et le joueur j , à mesure que celui-ci se sent moins proche de son partenaire, d'une période à l'autre.

En conséquence, l'agent i accorde, en $t + 1$, un poids inconditionnel d'autant plus faible aux gains de son partenaire, ce qui contribue à le désinciter à adopter un comportement prosocial, lors de cette période. Par symétrie, cette diminution de la valeur de δ_i agit comme une force poussant l'agent i à se comporter de manière plus égoïste, privilégiant ainsi son désir immédiat de gains matériels, au détriment d'une bonne entente (présente, et future) entre lui et son partenaire. Notre modèle est donc capable de prendre en compte le fait que les émotions négatives de colère et de déception, qui ont tendance à diminuer le bien-être psychique des individus, peuvent réduire le niveau de patience économique des agents. Par exemple, Lerner, Li et Weber (2012) ont montré qu'un état de tristesse exacerbe l'impatience financière des individus. A nouveau, le raisonnement que nous développons ici ne fait référence ni au concept d'escompte temporelle, ni à celui de préférence pour le présent. Cependant, il implique une certaine emprise des sentiments et des émotions dans le présent, qui peuvent induire des comportements ayant ensuite des conséquences dans le futur.

En outre, la diminution de δ_i implique que l'agent i dispose également, en $t + 1$, d'une capacité de résistance moindre face aux conséquences émotionnelles négatives entraînées par l'adoption d'un comportement inapproprié, de la part de j . Ce phénomène est illus-

tratif de la manière dont notre modèle intègre un autre aspect de l'influence dynamique délétère d'émotions négatives telles que la colère et la déception. Ces émotions conduisent en effet à augmenter la sensibilité des individus face à des événements négatifs, du fait qu'ils se sentent plus facilement offensés par un mauvais comportement (ce qui est sans doute en lien avec la capacité à faire preuve d'indulgence) et qu'ils se laissent plus facilement emporter par leur colère (McCullough, Bellah, Kilpatrick et Johnson, 2001).

Par ailleurs, il convient de noter ici que la réduction du capital de sympathie peut entraîner la négativité de ce paramètre. Lorsque c'est le cas, on désigne alors δ_i comme un capital de rancune/ressentiment, accumulé par l'agent i , vis-à-vis de l'agent j . Ce ressentiment témoigne du fait que la distance qui s'est installée entre l'agent i et son partenaire est telle, que celui-ci n'est plus seulement indifférent face à ce dernier ($\delta_i = 0$), mais en vient également à développer un désir inconditionnel de se venger ou d'éviter à l'avenir toute interaction avec lui, dès lors que cela est possible.

Finalement, l'accumulation d'un « capital affectif négatif », au sens où celui-ci est constitué sur la base de l'émergence d'émotions négatives, implique une certaine dette émotionnelle intertemporelle de l'agent j , qui devra être payée. Celle-ci peut alors être remboursée de deux façons. Premièrement, elle peut être « rachetée » par des actions prosociales futures (bénéficiant à i), de la part de j , ce dernier supportant alors le coût de la mise en œuvre de tels comportements. Sans cela, elle se paiera du fait des actions entreprises par l'agent i , sous l'influence de son état affectif, l'agent j supportant cette fois le coût d'être confronté au courroux de son partenaire.

6.1.1.1.3. Une évolution bornée

Les propriétés 3 et 4 servent à exprimer le fait que le capital de sympathie est borné et, plus précisément que $\forall i \in I$ et $\forall t \in (0, T)$, $\delta_i \in [-1, 1]$.

La justification de cet élément se rapporte, en partie, à ce dont nous avons déjà discuté lors du chapitre 5 (cf. 5.2.1.1.3). Nous avons précisé que, d'après les équations (5.2.3) et (5.2.4), lorsque le comportement du partenaire du joueur i , dans le jeu du bien public, respecte la valeur de la norme de contribution de ce dernier, soit $y_j = \bar{y}_i$, on obtient $\theta_{ij} = \delta_i$. Par conséquent, dans une telle situation, les poids que i accorde à ses propres gains matériels dans le jeu et à ceux du joueur j , sont respectivement égaux à 1 et δ_i .

Nous avons alors introduit l'idée qu'un niveau de capital de sympathie $\delta_i > 1$ semblait irréaliste car cela impliquerait que, même lorsque l'agent i est neutre émotionnellement face au comportement de son partenaire, il accorde plus de poids aux gains de ce dernier qu'il n'en accorde à ses propres gains.

De façon similaire, nous considérons que lorsque le partenaire du joueur i se comporte convenablement ($y_j = \bar{y}_i$), il semble également difficile d'envisager que l'agent i puisse avoir du ressentiment à son égard, au point d'accorder, aux gains de ce dernier, un poids négatif qui soit supérieur, en valeur absolue, au poids qu'il accorde à ses propres gains, ce qui nous conduit donc à supposer que $\delta_i \geq -1$.

Dans la prochaine partie, nous présentons comment l'évolution dynamique du capital de sympathie peut être généralisée aux versions du jeu où $n > 2$.

6.1.1.2. Le cas du groupe

Nous venons de voir comment les relations entre deux partenaires qui jouent au jeu du bien public évoluent en fonction des interactions qui se produisent à chaque répétition du jeu. Une question intéressante est alors de se demander comment cette dynamique est modifiée par l'introduction d'au moins un joueur supplémentaire.

On rappelle que, pour tout agent $i \in I$, $i \neq j$ et $j \in I$, on a :

$$\theta_{ij} = \delta_i + \varepsilon_{ij} \quad (6.1.5)$$

Il convient alors de remarquer que l'équation (6.1.4) implique que le même capital de sympathie est ressenti par l'agent i à l'encontre de chacun des autres membres du groupe.

Par ailleurs, lorsque les individus jouent en groupe, nous posons l'hypothèse que la grandeur qui conditionne l'évolution de son capital de sympathie correspond à un indice synthétique résumant l'évaluation qu'il fait, en moyenne, de l'adéquation morale du comportement de ses partenaires dans le jeu, pour la période en question (une telle hypothèse est ainsi en cohérence avec notre conceptions de la façon dont s'effectue l'évaluation d'autrui dans le modèle généralisé, en statique, cf. 5.3.1.1.1).

Ainsi, à chaque période t , l'évaluation globale et synthétique qu'effectue tout joueur i du comportement de l'ensemble de ses partenaires est défini, pour tout agent $i \in I$, $i \neq j$ et $j \in I$, par :

$$E_{ij,t}^* = \frac{1}{n-1} \sum_{j \neq i} \varepsilon_{ij,t}^* \quad (6.1.6)$$

Où on a toujours :

$$\varepsilon_{ij,t}^* = \alpha_i (y_{j,t}^* - \bar{y}_i)$$

Soit,

$$E_{ij,t}^* = \frac{1}{n-1} \sum_{j \neq i} \alpha_i (y_{j,t}^* - \bar{y}_i) \quad (6.1.7)$$

D'une période à la suivante, on suppose ensuite que c'est cette évaluation globale et synthétique qui fait évoluer le capital de sympathie initial, selon la relation suivante :

$$\delta_{i,t+1} = \delta_{i,t} + f(E_{ij,t}^*, \delta_{i,t}) \quad (6.1.8)$$

Où la fonction f est la même que celle qui est présentée en 6.1.1.1.

Le fait de supposer que le capital de sympathie s'applique invariablement selon les différents partenaires d'un même joueur est avantageux, puisque cela représente un moyen d'introduire les externalités, issues des comportements individuels de contribution, qui peuvent être observés au niveau du groupe.

En effet, notre spécification implique que lorsqu'un joueur entreprend un important effort contributif, il fait augmenter la moyenne de l'évaluation globale des comportements qui est déterminée par chacun de ses partenaires, ce qui tend ensuite à entraîner une augmentation de leur niveau de sympathie envers leurs partenaires et, puisque ce niveau de sympathie s'applique à l'ensemble des partenaires, cela renforce la proximité entre chacun des agents au sein du groupe.

Parallèlement, lorsqu'un joueur adopte un comportement de défection en contribuant faiblement, il fait diminuer la moyenne de l'évaluation globale des comportements qui est faite par chacun de ses partenaires, ce qui tend, toutes choses égales par ailleurs, à entraîner une diminution de leur niveau de sympathie et, puisque le capital de sympathie d'un agent s'applique à l'ensemble de ses partenaires, cela dégrade le niveau de proximité entre chacun des agents au sein du groupe.

6.1.1.3. Des agents myopes

Les éléments formels que nous avons vus jusqu'à maintenant dans ce chapitre décrivent comment le capital de sympathie varie d'une période à l'autre, en fonction de la manière dont les individus interagissent au cours des différentes répétitions du jeu. Pour exposer complètement la façon dont cette évolution dynamique se met en œuvre, il convient également de présenter le rapport que les individus entretiennent avec la temporalité.

De manière comparable au modèle de Van Dijk et Van Winden (1997), nous supposons ici que chaque agent est myope, au sens où il n'anticipe pas les conséquences potentielles de son comportement de contribution, sur la manière dont évolue le capital de sympathie que les autres agents développent (et qui s'applique à lui). Plus précisément, on suppose que ceux-ci ne sont pas conscients de l'impact que l'émergence, chez eux, des émotions morales d'évaluation d'autrui (sur la base des contributions réalisées lors d'une période), peut avoir sur l'évolution du niveau de sympathie qu'ils ressentent à l'égard de leurs partenaires (d'une période à l'autre). De même, les individus n'anticipent pas non plus l'influence que ces mêmes émotions peuvent avoir sur le développement du sentiment de proximité des autres joueurs, lorsque celles-ci émergent en eux. Finalement, la structure de décision peut ainsi être modélisée comme une séquence de décisions de contribution, qui ne sont connectées entre elles que par la variation se produisant, de période en période,

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

au niveau du capital de sympathie dont dispose chaque agent.

Ce qui nous démarque de l'approche de Van Dijk et Van Winden (1997), c'est que nous considérons, dans notre modèle, que lors de chaque prise de décision, les individus sont en mesure de prévoir l'émergence des émotions morales, aussi bien chez eux, que chez leurs partenaires, et que leur calcul économique tient compte des conséquences immédiates qui leurs sont attribuables, en termes de comportements réciproques (en particulier s'agissant des niveaux de récompenses). Ainsi, c'est plutôt la persistance dans le temps de tels états émotionnels, par l'intermédiaire de leur impact sur le capital de sympathie des joueurs, que les agents ne sont pas capables d'anticiper ou, dans l'hypothèse où ils en seraient capables, qu'ils choisissent d'ignorer lors de leur calcul.

Cette hypothèse nous semble pertinente dans la mesure où, dans la réalité, la capacité d'anticipation des agents, en ce qui concerne la persistance d'un état affectif chez autrui, est globalement limitée, notamment lorsqu'ils font face à des inconnus. Ainsi, bien que ceux-ci soient le plus souvent aptes à prévoir l'influence de leur comportement sur le ressenti émotionnel de leurs partenaires, à court terme, il est plus rare qu'ils parviennent naturellement à le faire à plus long terme. Bien entendu, certaines personnes font preuve d'une grande capacité d'anticipation émotionnelle et sont, à cet égard, en mesure d'adapter leur comportement stratégique, en tenant compte de la temporalité. Par exemple, les grands managers sportifs connaissent les subtilités inhérentes à la vie de groupe d'une équipe et sont capables d'encourager (ou à l'inverse de proscrire) certains comportements, chez leurs joueurs, car ils anticipent les conséquences bénéfiques (néfastes) que ceux-ci peuvent avoir, sur la dynamique des relations interpersonnelles au sein du groupe. Néanmoins, cette capacité fait partie des raisons pour lesquels ils occupent précisément ce poste, ainsi que des qualités à cultiver, par l'expérience, pour tenir le rôle spécifique qui leur incombe, au sein d'une équipe¹. En ce sens, une telle capacité ne semble pas être donnée à tout le monde et, en conséquence, notre approche se concentre plutôt sur l'analyse d'agents pour lesquels l'évolution dynamique des relations avec autrui, s'effectue principalement de manière inconsciente, sans qu'elle ne soit accessible à l'anticipation des agents.

Par ailleurs, lorsque les agents ne savent pas pendant combien de périodes ils seront amenés à interagir ensemble, il peut s'avérer très coûteux pour eux de réaliser une maximisation intertemporelle de leur bien-être, de surcroît en tenant compte de l'impact de leurs comportements sur leurs relations avec les autres membres de leur groupe. Dans ces conditions, il peut sembler rationnel pour les agents de se limiter au calcul de leur choix optimal, période par période, en ignorant la nature intertemporelle de leur interaction avec les membres de leur groupe d'appartenance actuel, bien que celle-ci ait finalement des conséquences dans l'évolution de leurs relations, dans les faits. De ce point de vue, la conception que nous employons pour modéliser la structure de décision des agents semble être adaptée à l'analyse d'agents placés dans des conditions d'incertitude en ce qui concerne la durée potentielle de leur interaction au sein d'un groupe.

1. Certains joueurs, comme le capitaine par exemple, peuvent également tenir un rôle spécifique au sein d'une équipe et mettre aussi leur expérience au service du collectif, de l'ambiance et de la bonne entente entre les partenaires, en montrant la voie et en agissant en « leader ».

Formellement, la résolution de notre modèle en dynamique implique donc de déterminer, à chaque période t , la contribution optimale de chaque agent, de la même façon que celle que nous avons présentée dans le chapitre 5. Puis, pour passer de la période t à la période $t + 1$, il convient ensuite de déterminer, pour chaque agent, l'évolution du capital de sympathie qui est entraînée par la réalisation effective de l'équilibre lors de la période t . Finalement, par cette procédure, on obtient la valeur de l'ensemble des paramètres de chaque agent pour la période $t + 1$, ce qui permet de calculer l'optimum réalisé en au cours de cette période, puis de déterminer l'évolution du capital de sympathie des agents entre $t + 1$ et $t + 2$ et ainsi de suite.

6.2. Simulations et résultats

Afin d'alléger la présentation de nos résultats, nous nous concentrons sur l'analyse de ce qui est obtenu pour des groupes composés de $n = 4$ joueurs.

Par ailleurs, on suppose, dans l'ensemble de nos simulations, que les groupes de joueurs restent inchangés pendant toute la durée du jeu et qu'ils interagissent ensemble durant $T = 15$ périodes. Comme suggéré plus haut, l'hypothèse de myopie des agents est plus réaliste vis-à-vis de participants placés dans des conditions d'incertitude en ce qui concerne la durée de l'interaction. En conséquence, on suppose que nos agents ne sont pas informés du nombre exact de périodes durant lesquelles ils vont jouer au moment de leur participation, ce qui est crédible au regard de démarches expérimentales comme celle de Rand *et al.* (2009).

Enfin, la dotation initiale individuelle de chaque agent est égale à $d = 10$, sauf lorsque cela est précisé, on fixe la valeur du rendement marginal du bien public à $a = 0.5$ et, dans le cas du JBP avec récompense, on suppose un coût unitaire lié à l'acte de récompenser qui est égal à $c = 0.5$.

Dans le cadre de ce chapitre, nous proposons d'analyser les résultats de notre modèle en observant d'une part ce que l'on obtient lorsque nous faisons jouer ensemble les profils de joueurs que nous avons utilisés dans le chapitre 5 (cf. 5.3.2.3) et d'autre part lorsque nous faisons jouer des joueurs dont les paramètres ont été tirés dans des distributions aléatoires, en suivant la même procédure que celle qui a été développée dans le chapitre 5 (cf. 5.3.2.3.2).

6.2.1. Résultats pour des profils de joueurs déterminés

Le tableau suivant rappelle la paramétrisation des quatre profils de joueurs que nous utilisons :

$i = \{1, 2, 3, 4\}$	δ_i	\bar{y}_i	β_i	α_i	ρ_i
Joueur 1	0.6	5	0.3	0.4	0.6
Joueur 2	0.4	4	0.5	0.5	0.5
Joueur 3	0.7	3	0.6	0.6	0.7
Joueur 4	0.5	4	0.4	0.5	0.4

TABLE 6.1.: Exemple de paramétrisation - Modèle dynamique

Tout d'abord, nous présentons les résultats que nous obtenons lorsque les joueurs participent, de manière répétée, au jeu du bien public standard.

6.2.1.1. Le JBP standard

Le graphique suivant indique, pour chaque période, les niveaux de contribution au JBP standard qui sont joués, à l'équilibre, par nos quatre joueurs :

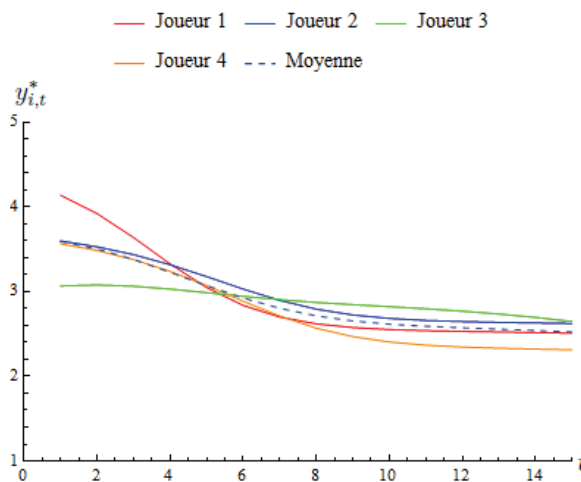


FIGURE 6.2.1.: Evolution des niveaux de contribution - JBP standard

On note que la courbe tracée en pointillés représente le niveau de contribution moyen observé pour chaque période.

Dans cet exemple illustratif, on constate une baisse tendancielle du niveau de coopération des agents au cours du temps. En moyenne, le niveau de contribution passe ainsi de 3.59 à la première période, à une valeur de 2.52 à la période 15.

Pour comprendre cette évolution dynamique des niveaux de contributions, il convient de s'intéresser à l'évolution dynamique du capital de sympathie des agents, qui est le seul paramètre à être modifié au cours du temps, dans notre modèle. Le graphique suivant représente, pour chaque période, le niveau de capital de sympathie caractérisant chacun des joueurs :

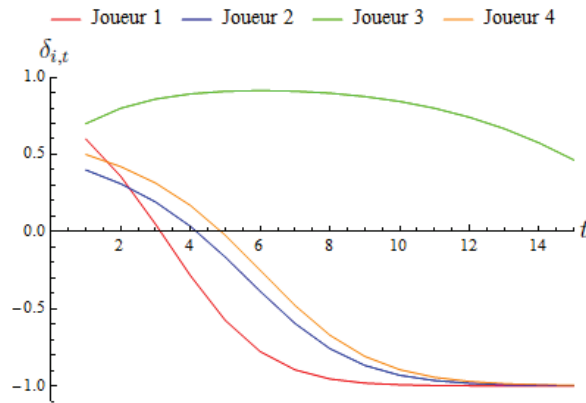


FIGURE 6.2.2.: Evolution du capital de sympathie - JBP standard

On constate, dans la figure 6.2.2, une forte chute du capital de sympathie des joueurs 1, 2 et 4. Celle-ci débute dès la première période et se prolonge ensuite sans discontinuité, de sorte que le capital de sympathie de ces joueurs tend vers sa borne inférieure qui est de -1 . Ainsi, ces trois joueurs ont, à la première période, été déçus et ont ressenti de la colère, du fait du comportement global qu'ils ont observé chez les autres membres du groupe. Ceci a provoqué une chute de leur capital de sympathie, avant d'entraîner une diminution de leur niveau de contribution lors de la période suivante, provoquant à nouveau l'émergence d'émotions négatives et ainsi de suite.

La déception et la colère ressenties par ces trois joueurs, lors de la première période, sont majoritairement déclenchées par la contribution relativement faible du joueur 3. Ce dernier, dont le comportement se rapproche le plus de celui d'un « passager clandestin » provoque, par sa contribution à la première période, l'émergence d'émotions négatives chez les autres, tandis que chez lui, c'est plutôt le sentiment de gratitude qui prédomine. Celui-ci est en effet relativement satisfait de pouvoir profiter de la contribution des autres membres de son groupe et cela entraîne une augmentation de son capital de sympathie lors des premières périodes du jeu. On voit ainsi comment, dans notre modèle, des asymétries dans l'évolution du capital de sympathie des joueurs sont susceptibles d'apparaître, ce qui est cohérent avec les résultats empiriques de Sonnemans *et al.* (2006), en ce qui concerne l'évolution des liens sociaux, dont nous avons discutés précédemment (cf. 4.2.2.). Cependant, à mesure que les sentiments négatifs des joueurs les plus coopératifs entraînent la diminution de leurs contributions, l'évaluation du joueur 3, à propos du comportement des autres individus de son groupe finit par déclencher l'émergence d'émotions négatives chez ce dernier aussi et on observe ainsi une diminution de son capital de sympathie lors de la deuxième moitié des répétitions du jeu.

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

Au total, cet exemple saisit la façon dont notre modèle prend en compte le fait que, dans le JBP standard, la présence d'un agent peu coopératif peut perturber l'évolution dynamique des relations au sein d'un groupe et comment l'émergence d'émotions négatives entre les partenaires peut conduire à une chute progressive du niveau global de coopération.

6.2.1.2. L'efficacité relative de la subvention

Imaginons maintenant que le planificateur social décide de promouvoir la coopération au sein de notre groupe de joueurs, dans le JBP standard, en utilisant le levier de la subvention, *i.e.* à travers une augmentation du rendement marginal du bien public. Toutefois, subventionner le bien public est coûteux et il pourrait être intéressant pour le planificateur de ne subventionner les contributions individuelles au bien public que sur un petit nombre de périodes. L'idée serait alors d'engager une dynamique positive entre les agents, au départ du jeu, à travers une coopération facilitée, pour ensuite profiter de la bonne entente créée entre les agents lors des périodes suivantes.

Supposons alors que le planificateur décide de subventionner les contributions individuelles de sorte à ce que le rendement marginal du bien public soit de $a = 0.75$, pendant les cinq premières périodes (soit un tiers de la durée de l'interaction)², avant que la valeur du rendement marginal ne s'établissent à $a = 0.5$, à partir de la période 6, et ce jusqu'à la fin du jeu.

Le graphique suivant représente les niveaux de contributions joués, à l'équilibre de chaque période, par notre groupe de quatre joueurs, dans la configuration que l'on vient de décrire :

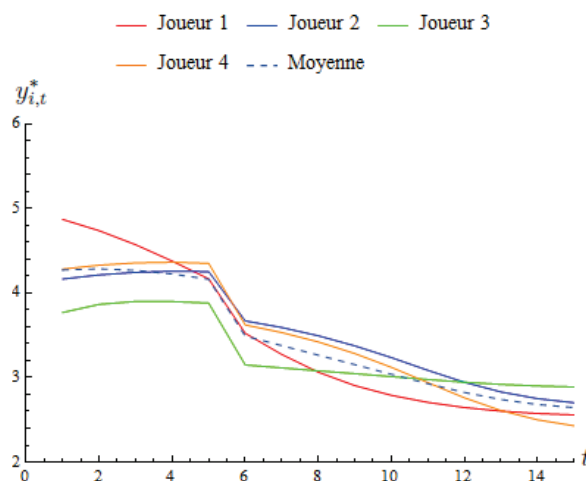


FIGURE 6.2.3.: Evolution des niveaux de contribution - Subvention sur 5 périodes

2. Les simulations montrent que l'on obtient des résultats similaires avec un nombre plus important de périodes où la subvention est effective et notamment lorsque l'on considère 8 périodes subventionnées, soit plus de la moitié de la durée de l'interaction.

Comme indiqué, la courbe tracée en pointillés désigne le niveau de contribution moyen observé pour chaque période.

Avec les profils de joueurs que nous utilisons dans cet exemple, on note que, dans cette configuration, la subvention permet de faire augmenter la contribution moyenne à la première période, en la faisant passer de 3.59 à 4.27. On remarque également que, mis à part le joueur 1, les autres individus ont tendance à augmenter leur niveau de contribution lors des cinq premières périodes, ce qui signifie que ces trois joueurs sont sous l'effet de la gratitude et que leur niveau de sympathie augmente pendant cette durée de jeu. En effet, on voit clairement un tel phénomène à l'observation du graphique suivant :

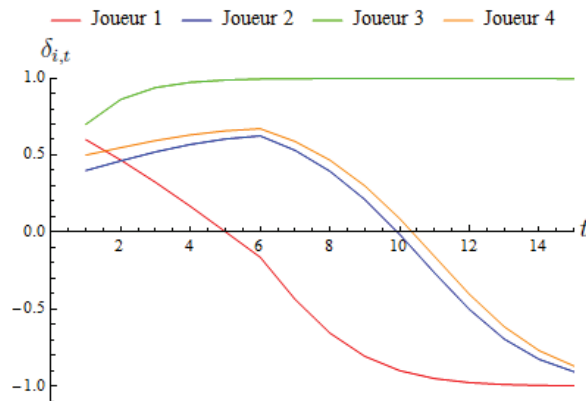


FIGURE 6.2.4.: Evolution du capital de sympathie - Subvention sur 5 périodes

Toutefois, on remarque que le même effet n'est pas obtenu pour le joueur 1 qui, malgré le léger effort coopératif consenti par ses partenaires, sous l'effet de la subvention, n'est pas complètement satisfait, et reste sous l'influence d'émotions négatives lors de l'évaluation du comportement de ces derniers. En conséquence, bien que les contributions des trois autres joueurs augmentent, la contribution moyenne diminue lors des cinq premières périodes, sous l'effet de la chute de la contribution du joueur le plus coopératif.

Par ailleurs, lors de l'arrêt de la subvention, on observe une chute de l'ensemble des contributions et, malgré l'évolution croissante du capital de sympathie des agents jusqu'alors (hors joueur 1), seul le capital de sympathie du joueur 3 continue à augmenter à partir de la période 6. Pour les autres agents, le niveau du capital de sympathie plonge vers la valeur -1 , en raison notamment des émotions négatives élicitées chez les joueurs 2 et 4, du fait de l'affaiblissement drastique du niveau de coopération, après l'arrêt de la subvention, ce qui se traduit par une tendance à la baisse prononcée des contributions, jusqu'à la fin du jeu. Finalement, l'incitation économique de la subvention n'a pas été suffisante, compte tenu du profil des joueurs, pour instaurer une dynamique qui soit suffisamment positive pour limiter la diminution des contributions après son arrêt. Son effet principal a plutôt été de ralentir cette diminution pendant la durée de sa mise en application.

Maintenant, intéressons nous à ce qui se passe lorsque l'on fait participer ces mêmes joueurs à un JBP avec phase de récompense.

6.2.1.3. L'efficacité de l'ajout d'une phase de récompense au JBP

Le graphique qui suit représente, pour chaque période, les niveaux de contribution joués, à l'équilibre, par nos quatre joueurs, dans le JBP avec récompense :

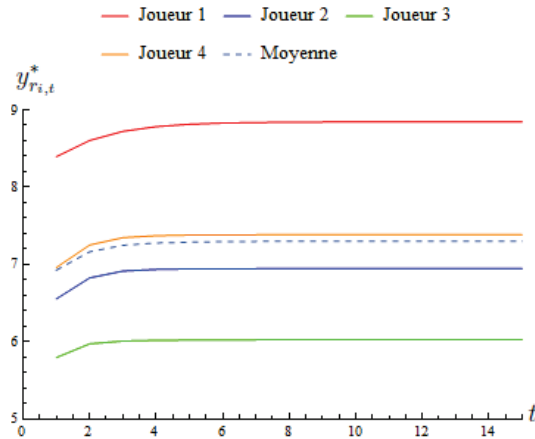


FIGURE 6.2.5.: Evolution des niveaux de contribution - JBP récompense

Il convient à nouveau de noter que la courbe tracée en pointillés représente le niveau de contribution moyen observé pour chaque période.

Dans cet exemple illustratif, on constate que, contrairement au cas du JBP standard (qu'il y ait ou non une subvention lors des premières périodes), la coopération ne s'effondre pas au cours du temps dans le JBP avec récompense. Au contraire, le niveau de contribution moyen démontre une légère augmentation, puisqu'il s'établit à 6.92 à la première période, avant de se stabiliser sur la valeur 7.29, à l'approche des dernières périodes. Ainsi, tout l'intérêt de l'introduction d'une phase de récompense, au jeu du bien public, est que celle-ci est capable de susciter un niveau de coopération nettement supérieur à celui qui est obtenu en son absence, lors de la première période, et que cela permet ensuite à la coopération de se stabiliser dynamiquement, sous l'effet du rapprochement social que cela génère entre les partenaires.

En effet, cette évolution dynamique des niveaux de contributions est à nouveau à imputer à l'évolution dynamique du capital de sympathie des agents. Le graphique suivant représente, pour chaque période, le niveau de capital de sympathie qui caractérise chaque joueur dans cette version du jeu :

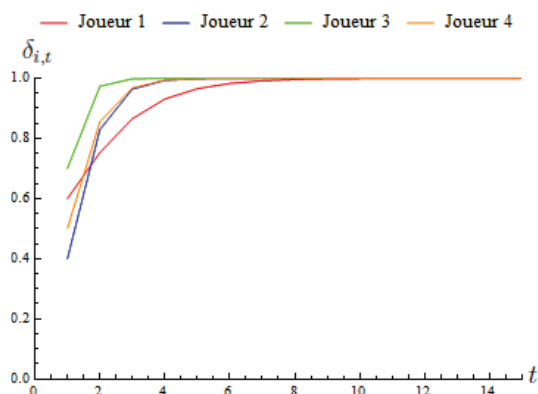


FIGURE 6.2.6.: Evolution du capital de sympathie - JBP récompense

A l'observation de la figure 6.2.6, on remarque que, dans le cas du JBP avec récompense, le capital de sympathie de l'ensemble des joueurs connaît une forte augmentation, dès les premières périodes du jeu, pour ensuite converger vers sa borne supérieure qui est de 1. Cette évolution favorable provient du fait que, malgré des niveaux de contributions différents (notamment en ce qui concerne le joueur 3, qui contribue le moins), l'ensemble des joueurs coopère suffisamment pour faire émerger de la gratitude chez chacun des partenaires. En témoigne les niveaux de récompenses versés et reçus par chaque joueur, que l'on retrouve dans le graphique suivant :

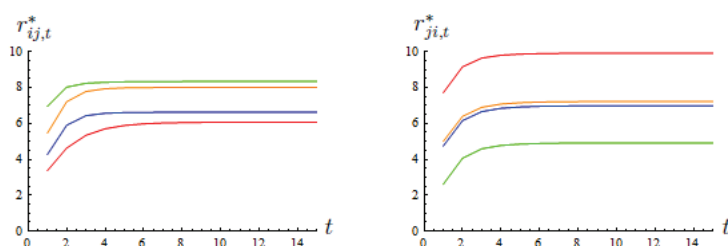


FIGURE 6.2.7.: Evolution des niveaux de récompense versés et reçus

Ainsi, sur la base des sentiments positifs de gratitudes qui émergent, dans cette version du jeu, du fait de l'important niveau de coopération obtenu grâce au mécanisme incitatif de récompense, les individus développent de la sympathie les uns pour les autres, se sentent plus proches les uns des autres, et leurs bonnes relations au sein du groupe permettent le maintien et même le renforcement de l'effort coopératif au cours du temps.

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

En comparaison avec le levier de la subvention, seule l'interaction ciblée que constitue la phase récompense est en mesure à la fois de promouvoir le niveau de coopération, mais également de le « protéger » à long terme, en orientant les agents vers des comportements favorisant l'émergence de sentiments positifs tels que la gratitude et la fierté, plutôt que vers ceux de colère ou de déception, lesquels sont généralement suscités dans le JBP standard.

Au total, en regroupant les mêmes individus qui n'ont pas réussi à coopérer dans le JBP standard, au sein d'un contexte où ils ont l'opportunité d'avoir une interaction ciblée et positive avec chacun de leurs partenaires, ceux-ci sont maintenant capables de développer de manière pérenne de bonnes relations interpersonnelles, ainsi que de maintenir un effort coopératif qui se rapproche de l'optimum social du jeu, grâce à l'émergence d'émotions positives telles que la gratitude et la fierté.

Dans le but d'examiner la portée générale des résultats que nous avons obtenus dans cet exemple spécifique, nous proposons, dans la section suivante, l'analyse des résultats de notre modèle, lorsque les joueurs sont générés aléatoirement.

6.2.2. Résultats pour des profils de joueurs générés aléatoirement

En utilisant la même procédure que dans la section 5.3.2.3.2 pour générer aléatoirement le profil de chacun des joueurs, on fait participer cent groupes de $n = 4$ agents aux deux versions du jeu. Les deux graphiques suivants donnent alors l'évolution moyenne des niveaux de contribution et des niveaux de capital de sympathie que l'on obtient dans le JBP standard :

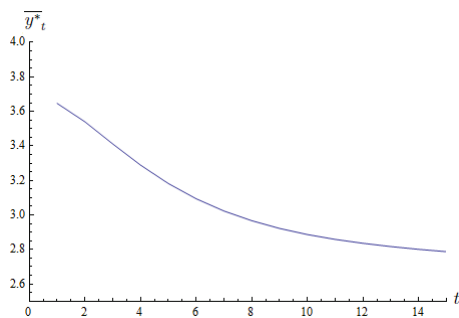


FIGURE 6.2.8.: Evolution du niveau moyen de contribution - JBP standard

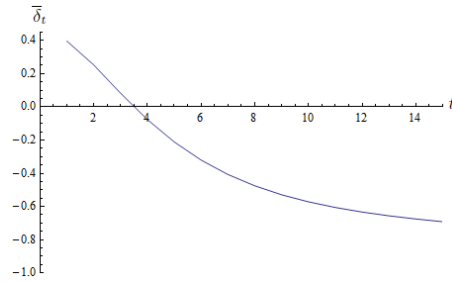


FIGURE 6.2.9.: Evolution du capital de sympathie moyen - JBP standard

D'après ce que l'on observe dans les graphiques 6.2.8 et 6.2.9, on constate que les résultats obtenus avec des groupes aléatoirement générés sont conformes avec ceux que nous avons présentés de manière détaillée dans l'exemple illustratif, ce qui tend à démontrer la pertinence, ainsi que la généralité des interprétations que nous avons pu en faire. En effet, on observe à nouveau une tendance à la baisse des niveaux moyens de contribution et de capital de sympathie, ce qui témoigne de l'incapacité du JBP standard à susciter suffisamment de coopération pour permettre aux individus de créer et d'entretenir de bonnes relations interpersonnelles (tout du moins pour ce qui est de populations d'agents correspondant à nos hypothèses).

Ces résultats contrastent avec ce que l'on obtient lorsque l'on fait participer cent groupes de $n = 4$ joueurs au JBP avec récompense :

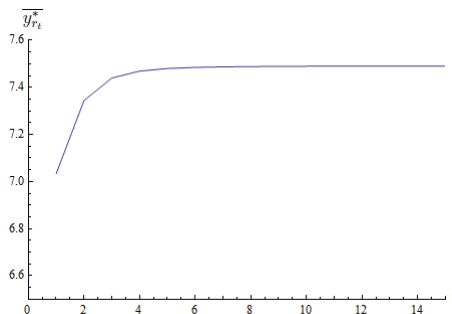


FIGURE 6.2.10.: Evolution du niveau moyen de contribution - JBP récompense

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

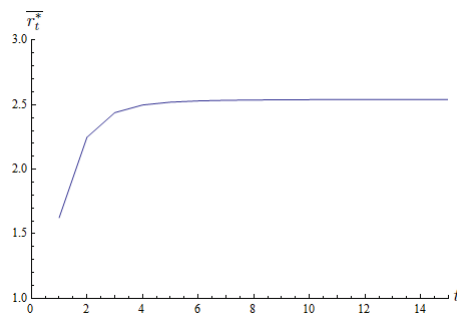


FIGURE 6.2.11.: Evolution du niveau moyen de récompense (versé) - JBP récompense

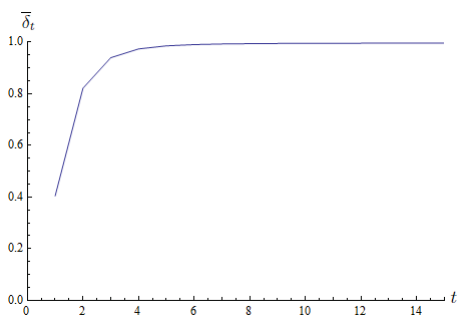


FIGURE 6.2.12.: Evolution du capital de sympathie moyen - JBP récompense

En effet, il apparaît clairement dans nos résultats que, lorsqu'est introduite l'option de récompense, celle-ci est largement utilisée par les agents, sous l'effet de l'émotion de gratitude, et que, par anticipation de ce gain supplémentaire, et de l'approbation morale qu'il communique (et ainsi du sentiment de fierté qu'il élicite), un niveau conséquent de coopération est suscité au sein des groupes lors de la première période. Ces conditions permettent alors, sous l'influence de l'émotion de gratitude, d'augmenter la sympathie qu'ont les individus les uns envers les autres, ainsi que leur proximité sociale et que le lien qui les unit, ce qui favorise la création, le maintien, et le renforcement de bonnes relations interpersonnelles entre les agents. En conséquence, les résultats de notre modèle suggèrent que l'introduction d'une phase de récompense, dans le jeu du bien public, est un moyen efficace de promouvoir la coopération et la bonne entente entre les agents, dans des groupes où les individus interagissent de manière répétée au sein d'un même groupe.

6.3. Discussion

La démarche formelle que nous avons suivie dans ce travail de thèse s'est inscrite dans le projet ambitieux de représenter comment l'influence d'un ensemble d'émotions morales clés peut, dans la situation de dilemme social que constitue le jeu du bien public, expliquer l'évolution dans le temps des comportements et des relations interpersonnelles au sein d'un groupe. Le défi était alors de représenter le plus simplement possible, l'immense complexité que représente le déroulement d'une interaction répétée entre plusieurs individus. En particulier, il s'agissait de prendre en compte l'hétérogénéité de différents types de personnalité que l'on retrouve dans les populations humaines, ainsi que la pluralité de réactions qui peut résulter de la rencontre entre ces individualités. Ceci nous a conduit à développer un modèle détaillé, possédant de nombreux paramètres et variables endogènes, et dont le niveau de raffinement permet d'englober une large variété d'éléments prenant part à la décision économique dans les dilemmes sociaux. Les éléments centraux caractéristiques de ce modèle sont les suivants : (1) la prise en compte de la pluralité des affects (à valence négative et positive) et de leur impact sur la décision, dans un cadre compatible avec les théories modernes et contemporaines de l'émotion, (2) l'introduction d'une phase de récompense, lors du processus de coopération, et des mécanismes affectifs (comme la gratitude) qui y sont associés et (3) l'ajout d'une dynamique émotionnelle et relationnelle basée sur l'histoire du jeu et sur les comportements passés des joueurs.

Nous abordons maintenant certains prolongements envisageables à ce travail de modélisation, qui représentent autant de pistes vers des extensions futures visant à en renforcer la pertinence.

Tout d'abord, nous supposons dans notre modèle que chaque individu a la même dotation initiale avant de débiter chaque période du jeu. Toutefois, il convient de se demander comment notre modèle pourrait être en mesure d'analyser la décision d'agents n'ayant pas les mêmes dotations initiales. Deux cas nous semblent alors pertinents à étudier. Premièrement, le cas où chaque agent est informé de la diversité des dotations et en connaît la répartition au sein de son groupe. Dans une telle situation, on peut imaginer que la valeur de la norme de contribution utilisée par chaque agent, pour analyser la moralité de sa propre conduite, dépend de sa dotation initiale individuelle ($\bar{y}_i = p_i \times d_i$), tandis que celle qui est utilisée pour évaluer la bienveillance de ses partenaires s'établit (ou non, selon le degré d'empathie de l'individu) en fonction de leurs dotations initiales particulières ($\bar{y}_{ij} = p_i \times d_j$). Deuxièmement, on peut aussi supposer que chaque agent ignore que ses partenaires ont potentiellement des dotations initiales différentes de la sienne, auquel cas la capacité d'empathie de chaque agent est mécaniquement biaisée et on peut imaginer que cela résulte en une plus grande difficulté de mise en place de la coopération, étant donné l'incompatibilité des points de vue des différents membres du groupe. Cette hypothèse mériterait à notre sens d'être testée dans un modèle étendu.

Par ailleurs, dans notre modèle, l'information est supposée complète et parfaite, ce qui constitue une hypothèse forte et assez peu réaliste. Celle-ci implique que les individus sont en mesure d'identifier et de saisir complètement la personnalité de leurs partenaires, leurs conceptions de la norme de comportement moralement adapté et l'intensité avec laquelle leurs émotions vont émerger. Ceux-ci sont ainsi parfaitement en mesure d'anticiper la réaction de chacun des membres de leur groupe, en conséquence de leur propre décision.

6. L'évolution dynamique des relations entre agents dans le jeu du bien public

Ainsi, une possibilité pour rendre notre modèle plus réaliste, serait d'intégrer, chez les agents, un certain niveau d'incertitude vis-à-vis de la personnalité de leurs partenaires. Une première possibilité, la plus simple selon nous, pourrait être d'introduire de l'incertitude vis-à-vis de ce qui conditionne le déclenchement de l'émergence des émotions chez autrui, soit, dans notre modèle, la valeur de sa norme de contribution. Plus généralement, tout ce qui participe à définir la personnalité des agents, dans le modèle, soit le niveau de l'intensité de l'émergence de leurs émotions, est susceptible d'être envisagé comme une information uniquement disponible à l'individu.

Une autre hypothèse forte que nous posons dans notre formalisation concerne la façon dont les émotions d'évaluation d'autrui émergent. En particulier, notre spécification suppose implicitement que les différentes émotions d'évaluation d'autrui qui sont susceptibles d'être élicitées chez un agent, émergent avec la même intensité (désignée par le paramètre α), quelle qu'en soit leur valence. Il n'y a pas de raison que cela soit le cas et, selon toute vraisemblance, il est même plus réaliste de supposer le contraire. D'après la théorie des émotions, en effet, le ressenti émotionnel (son intensité en particulier) présente une dimension personnelle et contextuelle qui dépend de la nature de l'émotion en jeu (sa valence, son type, sa dimension cognitive etc.). Techniquement, une façon simple de contourner cette limite aurait été de recourir à des fonctions définies par parties, ce qui compliquerait toutefois la résolution analytique du modèle et conduirait également à utiliser un paramètre supplémentaire.

En outre, un autre prolongement envisageable serait de définir, pour chaque agent i , un capital de sympathie affecté individuellement à chacun de ses partenaires j (noté δ_{ij}). Ceci permettrait, en complément du capital de sympathie δ_i , dont l'évolution est relative au comportement moyen observé dans le reste du groupe, de mieux prendre en compte le fait que chaque individu retient les comportements particuliers de ses différents partenaires et que cela peut influencer ses choix, notamment en ce qui concerne les niveaux de récompenses qu'il leur offre. En effet, ceci pourrait jouer un rôle important lorsque l'on considère un nombre réduit de partenaires, qui apprennent à se connaître à mesure que les interactions se produisent dans le jeu. La portée générale de notre modèle nous a conduit à nous placer dans l'optique d'un capital de sympathie de chaque agent qui s'applique indifféremment à l'ensemble des partenaires, car cela nous semble plus compatible avec ce qui se produit à grande échelle. Toutefois, s'agissant de groupes de taille réduite, l'introduction d'une telle nuance pourrait être en mesure d'appréhender les affinités pouvant se créer entre certains membres d'une équipe et donnant éventuellement lieu à l'apparition de « sous-groupes ».

Enfin, empiriquement, certaines données suggèrent que les individus ont tendance à considérer la contribution moyenne du groupe comme le point de référence à partir duquel évaluer le comportement de leurs partenaires et, possiblement, comme un indice de la norme collective qui se dégage du groupe spécifique dans lequel ils sont intégrés. Ceci pose la question de la capacité de mise à jour de la norme de contribution intériorisée par chaque agent et soulève également la question d'une préférence pour le fait de se coordonner sur un comportement commun, ce qui relèverait d'une certaine forme de conformisme. Dans le prolongement de notre approche dynamique, on pourrait supposer que les individus disposent d'une telle capacité d'apprentissage de la « culture » qui émerge de leur propre groupe, en ce qui concerne les comportements de contribution. Formellement, ceci

impliquerait alors une évolution du niveau de la norme de contribution des agents, en fonction du résultat de l'interaction ayant lieu lors de chaque période (ou éventuellement seulement lors des premières). Dans cette perspective, on peut imaginer que l'efficacité du mécanisme incitatif de récompense n'en soit que renforcée car, étant donné qu'il permet une contribution moyenne relativement élevée à la première période, cela pourrait alors orienter les groupes d'agents vers l'adoption d'une norme de contribution spécifique, commune à l'ensemble des partenaires, qui évolue à la hausse au cours du temps.

Cinquième partie .
Conclusion Générale

Conclusion générale

Dans ce travail de thèse, nous avons étudié le rôle moteur des émotions morales dans la prise de décision des agents, lorsqu'ils sont placés dans une situation de dilemme social.

Dans le premier chapitre, notre revue de littérature a permis de mettre en perspective des résultats issus de l'économie expérimentale et de la psychologie sociale, qui soulignent l'importance des sentiments moraux dans le comportement effectif des individus, lorsqu'ils rentrent en interaction avec autrui. Ainsi, nous avons mis en évidence que les émotions morales à valence négative ont la capacité d'influencer les comportements stratégiques des individus en les plaçant dans une optique de long terme. Lors de leur prise de décision, l'anticipation de l'émergence d'émotions telles que la culpabilité ou la honte s'oppose en effet à la recherche de l'intérêt personnel immédiat en laissant augurer les conséquences émotionnelles négatives de comportements opportunistes, ce qui agit en faveur de comportements prosociaux comme la coopération. Notre revue souligne cependant les limites des émotions morales à valence négative en termes de bien-être et parfois d'efficacité à promouvoir de tels comportements, ce qui nous a conduit à nous intéresser particulièrement aux émotions morales à valence positive, comme la gratitude, qui ont été négligées dans la littérature en économie.

Dans le deuxième chapitre, nos éléments de revue ont ainsi fait ressortir le rôle essentiel de la gratitude dans les interactions sociales, que ce soit à travers son implication dans la création et le renforcement des relations interpersonnelles, son influence dans l'adoption de comportements prosociaux tels que la coopération, ou pour son impact sur le bien-être, l'efficacité ainsi que la motivation des agents. Ces conclusions nous ont conduit à prendre la mesure de la nécessité, pour les économistes, de ne pas négliger le rôle de cette émotion, s'agissant notamment du contexte des dilemmes sociaux, et par conséquent de chercher à l'intégrer, avec précision, dans leur représentation du processus de décision des agents.

Le chapitre 3 nous a ensuite permis de retracer le parcours suivi par les économistes comportementalistes qui ont progressivement accordé, dans leurs travaux de modélisation, de plus en plus de place à l'influence des normes, des intentions, des attentes et, en filigrane, au rôle des émotions (morales). Tout en restant dans le cadre de la rationalité, l'aménagement du modèle classique, en théorie des jeux, à travers le relâchement de l'hypothèse de la recherche de l'intérêt individuel et l'intégration de ces différents mécanismes psychologiques, a abouti à la création de modèles de préférences sociales bénéficiant d'une représentation plus fidèle des êtres humains et de leur manière de fonctionner en société. Ainsi, l'intégration précise, explicite et systématique du rôle des émotions morales nous est apparue comme une voie permettant d'accroître encore davantage le niveau de réalisme et de capacité prédictive des modèles d'économie comportementale.

L'objet principal du chapitre 4 a ensuite été de définir le cadre conceptuel dans lequel réaliser notre contribution originale, de façon à prendre précisément en compte les spécificités de l'influence particulière de la gratitude, ainsi que celle d'autres émotions morales comme la colère ou la culpabilité. Le modèle se focalise sur le jeu du bien public qui constitue un cadre riche, par la variété de ses déclinaisons, mais aussi par les nombreuses études empiriques qui lui sont consacrées. Une revue non exhaustive des résultats expérimentaux relatifs aux différentes versions du jeu du bien public nous a ainsi permis de cibler les différents déterminants de la coopération dans ce jeu. Enfin, nous avons également détaillé les raisons qui font que le cadre de ce jeu convient idéalement à notre analyse, comme par exemple le fait qu'il touche à de nombreuses situations économiques concrètes (abstentionnisme, évasion fiscale, paiement des transports, etc.) et concerne de nombreux domaines d'investigation (l'environnement, la santé, la gestion des organisations etc.). Le jeu du bien public peut se jouer en groupe, ce qui alimente les interactions entre joueurs, a vocation à être répété, ce qui autorise la mise en place d'une dynamique comportementale affective, et, surtout, la nature du jeu fait que la plupart des émotions morales peuvent être suscitées, chez les agents, lors de leur participation.

Enfin, nous proposons, dans les chapitres 5 et 6, un travail de formalisation original qui a pour but d'expliquer et d'analyser, en statique et en dynamique comment l'ajout d'une phase de récompense, au jeu du bien public, est en mesure de promouvoir la coopération entre les joueurs.

Au total, les résultats de notre modèle suggèrent que le mécanisme incitatif de récompense est efficace pour promouvoir la coopération des agents dans le jeu du bien public joué en un coup, comparativement à ce qui est obtenu en son absence.

Dans la version standard du jeu du bien public, le manque d'incitation économique à coopérer tend en effet à entraîner l'émergence d'émotions négatives, lorsque les agents évaluent le comportement de leurs partenaires, ce qui les désincite à contribuer, notamment en raison de mécanismes de réciprocité négative. Bien que l'augmentation du rendement marginal par la subvention soit une solution permettant de limiter le manque de coopération dans le jeu, son efficacité est relativement limitée en comparaison avec l'effet qu'entraîne l'introduction d'une phase subséquente de récompense.

En effet, lorsque l'option de récompense est disponible, elle offre aux agents la possibilité d'avoir une interaction ciblée avec chacun de leurs partenaires. Cette dernière sert alors de canal de communication aux individus pour exprimer à autrui la gratitude qui émerge en eux, sous la forme d'une récompense monétaire (dans notre version du jeu). Dès lors, la potentialité d'être récompensé représente une incitation économique supplémentaire à contribuer, que ce soit pour des considérations d'ordre matériel ou pour obtenir l'approbation de nos partenaires, ce qui permet de susciter une plus grande coopération dans le jeu.

Notre analyse du jeu en dynamique a par la suite permis de mettre en évidence que, sur la base des sentiments positifs de gratitude qui émergent, dans le jeu du bien public avec phase de récompense, les individus développent de la sympathie, au cours du temps, et deviennent plus proches les uns des autres. Comparativement à la version standard du jeu, où les émotions à valence négative de colère et de déception dominent et conduisent à une dégradation des relations interpersonnelles, les résultats de la modélisation nous

enseignent que l'ajout d'une phase de récompense favorise le développement de bonnes relations au sein des groupes, ainsi que le maintien (voire le renforcement) de l'effort coopératif au cours du temps. De plus, notre modèle suggère également que, comparativement au levier de la subvention, seule l'interaction ciblée que constitue la phase de récompense est en mesure à la fois de promouvoir efficacement le niveau de coopération et d'empêcher sa diminution à long terme.

Qu'il s'agisse du cas statique ou du cas dynamique, les résultats offerts par notre modélisation sont d'autant plus convaincants, qu'ils bénéficient globalement d'une stabilité manifeste, notamment face à la question primordiale de l'hétérogénéité des agents.

Pour conclure, ce travail présente un modèle novateur et intéressant puisque, conformément à la littérature qui concerne la théorie des émotions, il envisage les émotions morales sous la forme d'outils d'aide à la décision qui ont tous (parfois simultanément, le plus souvent tour à tour, selon le contexte) une fonction importante dans le choix final des agents. En particulier, ce travail de formalisation insiste sur la part centrale que peuvent avoir les émotions positives (comme la gratitude, la fierté, l'admiration, la joie etc.), en lien avec la dynamique relationnelle, dans l'émergence des comportements prosociaux tels que la coopération.

Bibliographie

- [1] G. A. AKERLOF, *A theory of social custom, of which unemployment may be one consequence*, The quarterly journal of economics, (1980), pp. 749–775.
- [2] S. B. ALGOE, *Find, remind, and bind : The functions of gratitude in everyday relationships*, Social and Personality Psychology Compass, 6 (2012), pp. 455–469.
- [3] S. B. ALGOE, S. L. GABLE, AND N. C. MAISEL, *It's the little things : Everyday gratitude as a booster shot for romantic relationships*, Personal relationships, 17 (2010), pp. 217–233.
- [4] S. B. ALGOE, J. HAIDT, AND S. L. GABLE, *Beyond reciprocity : gratitude and relationships in everyday life.*, Emotion, 8 (2008), pp. 425–429.
- [5] M. D. ALICKE, D. A. DUNNING, AND J. KRUEGER, *The self in social judgment*, Psychology Press, 2005.
- [6] J. A. AMEGASHIE, *Economics, gratitude, and warm glow*, tech. report, University of Guelph, Department of Economics and Finance, 2006.
- [7] J. ANDREONI, *Why free ride ? : Strategies and learning in public goods experiments*, Journal of public Economics, 37 (1988), pp. 291–304.
- [8] —, *Giving with impure altruism : Applications to charity and ricardian equivalence*, The Journal of Political Economy, 97 (1989), pp. 1447–1458.
- [9] —, *Impure altruism and donations to public goods : A theory of warm-glow giving*, The economic journal, 100 (1990), pp. 464–477.
- [10] —, *Cooperation in public-goods experiments : kindness or confusion ?*, The American Economic Review, 85 (1995), pp. 891–904.
- [11] J. ANDREONI AND R. PETRIE, *Public goods experiments without confidentiality : a glimpse into fund-raising*, Journal of Public Economics, 88 (2004), pp. 1605–1623.
- [12] K. ARROW, *Political and economic evaluation of social effects and externalities*, in The analysis of public output, NBER, 1970, pp. 1–30.
- [13] D. BALLIET, *Communication and cooperation in social dilemmas : A meta-analytic review*, Journal of Conflict Resolution, 54 (2009), pp. 39–57.
- [14] D. BALLIET, L. B. MULDER, AND P. A. VAN LANGE, *Reward, punishment, and cooperation : a meta-analysis.*, Psychological bulletin, 137 (2011), pp. 594–615.
- [15] A. BARR, *Social dilemmas, shame-based sanctions, and shamelessness : experimental results from rural zimbabwe*, Centre for the Study of African Economies Working Paper WPS/2001.11 : Oxford University, (2001).
- [16] S. G. BARSADE AND D. E. GIBSON, *Why does affect matter in organizations ?*, The Academy of Management Perspectives, 21 (2007), pp. 36–59.
- [17] M. Y. BARTLETT, P. CONDON, J. CRUZ, J. BAUMANN, AND D. DESTENO, *Gratitude : Prompting behaviours that build relationships*, Cognition & emotion, 26 (2012), pp. 2–13.

- [18] M. Y. BARTLETT AND D. DESTENO, *Gratitude and prosocial behavior helping when it costs you*, *Psychological science*, 17 (2006), pp. 319–325.
- [19] C. D. BATSON, *Moral masquerades : Experimental exploration of the nature of moral motivation*, *Phenomenology and the Cognitive Sciences*, 7 (2008), pp. 51–66.
- [20] C. D. BATSON, E. COLLINS, AND A. A. POWELL, *Doing business after the fall : The virtue of moral hypocrisy*, *Journal of Business Ethics*, 66 (2006), pp. 321–335.
- [21] C. D. BATSON, J. FULTZ, AND P. A. SCHOENRADE, *Distress and empathy : Two qualitatively distinct vicarious emotions with different motivational consequences*, *Journal of personality*, 55 (1987), pp. 19–39.
- [22] C. D. BATSON AND L. L. SHAW, *Evidence for altruism : Toward a pluralism of prosocial motives*, *Psychological inquiry*, 2 (1991), pp. 107–122.
- [23] C. D. BATSON AND E. R. THOMPSON, *Why don't moral people act morally ? motivational considerations*, *Current directions in psychological science*, 10 (2001), pp. 54–57.
- [24] P. BATTIGALLI AND M. DUFWENBERG, *Guilt in games*, *The American economic review*, 97 (2007), pp. 170–176.
- [25] R. F. BAUMEISTER, E. BRATSLAVSKY, C. FINKENAUER, AND K. D. VOHS, *Bad is stronger than good.*, *Review of general psychology*, 5 (2001), pp. 323–370.
- [26] G. BEN-SHAKHAR, G. BORNSTEIN, A. HOPFENSITZ, AND F. VAN WINDEN, *Reciprocity and emotions in bargaining using physiological and self-report measures*, *Journal of economic psychology*, 28 (2007), pp. 314–323.
- [27] R. BÉNABOU AND J. TIROLE, *Incentives and prosocial behavior*, *The American economic review*, 96 (2006), pp. 1652–1678.
- [28] S. BLOUNT, *When social outcomes aren't fair : The effect of causal attributions on preferences*, *Organizational behavior and human decision processes*, 63 (1995), pp. 131–144.
- [29] G. E. BOLTON AND A. OCKENFELS, *Erc : A theory of equity, reciprocity, and competition*, *American economic review*, (2000), pp. 166–193.
- [30] R. BOSMAN AND F. VAN WINDEN, *Emotional hazard in a power-to-take experiment*, *The Economic Journal*, 112 (2002), pp. 147–169.
- [31] S. BOWLES AND H. GINTIS, *Prosocial emotions*, New York : Oxford University Press, 2005, pp. 339–366.
- [32] K. A. BREKKE, K. E. HAUGE, J. T. LIND, AND K. NYBORG, *Playing with the good guys. a public good game with endogenous group formation*, *Journal of Public Economics*, 95 (2011), pp. 1111–1118.
- [33] K. A. BREKKE, S. KVERNDOKK, AND K. NYBORG, *An economic model of moral motivation*, *Journal of public economics*, 87 (2003), pp. 1967–1983.
- [34] J. BROSIG, J. WEIMANN, AND A. OCKENFELS, *The effect of communication media on cooperation*, *German Economic Review*, 4 (2003), pp. 217–241.
- [35] B. R. BROWN, *Face-saving following experimentally induced embarrassment*, *Journal of Experimental Social Psychology*, 6 (1970), pp. 255–271.
- [36] M. N. BURTON-CHELLEW AND S. A. WEST, *Prosocial preferences do not explain human cooperation in public-goods games*, *Proceedings of the National Academy of Sciences*, 110 (2013), pp. 216–221.

- [37] C. CAMERER, *Behavioral game theory : Experiments in strategic interaction*, Princeton University Press, 2003.
- [38] C. CAMERER, G. LOEWENSTEIN, AND D. PRELEC, *Neuroeconomics : How neuroscience can inform economics*, Journal of economic Literature, 43 (2005), pp. 9–64.
- [39] L. A. CAMERON, *Raising the stakes in the ultimatum game : Experimental evidence from indonesia*, Economic Inquiry, 37 (1999), pp. 47–59.
- [40] J. P. CARPENTER, A. G. DANIERE, AND L. M. TAKAHASHI, *Cooperation, trust, and social capital in southeast asian urban slums*, Journal of Economic Behavior & Organization, 55 (2004), pp. 533–551.
- [41] J. P. CARPENTER AND P. H. MATTHEWS, *Norm enforcement : anger, indignation, or reciprocity ?*, Journal of the European Economic Association, 10 (2012), pp. 555–572.
- [42] G. CHARNESS AND B. GROSSKOPF, *Relative payoffs and happiness : an experimental study*, Journal of Economic Behavior & Organization, 45 (2001), pp. 301–328.
- [43] G. CHARNESS AND E. HARUVY, *Altruism, equity, and reciprocity in a gift-exchange experiment : an encompassing approach*, Games and Economic Behavior, 40 (2002), pp. 203–231.
- [44] G. CHARNESS AND M. RABIN, *Understanding social preferences with simple tests*, Quarterly journal of Economics, (2002), pp. 817–869.
- [45] A. CHAUDHURI, *Sustaining cooperation in laboratory public goods experiments : a selective survey of the literature*, Experimental Economics, 14 (2011), pp. 47–83.
- [46] L. H. CHEN, M.-Y. CHEN, AND Y.-M. TSAI, *Does gratitude always work ? ambivalence over emotional expression inhibits the beneficial effect of gratitude on well-being*, International Journal of Psychology, 47 (2012), pp. 381–392.
- [47] L. H. CHEN AND Y. H. KEE, *Gratitude and adolescent athletes' well-being*, Social Indicators Research, 89 (2008), pp. 361–373.
- [48] L. H. CHEN, C.-H. WU, AND S. CHEN, *Gratitude and athletes' life satisfaction : a intra-individual analysis on the moderation of ambivalence over emotional expression*, Social Indicators Research, 123 (2015), pp. 227–239.
- [49] J.-K. CHOI AND T. AHN, *Strategic reward and altruistic punishment support cooperation in a public goods game experiment*, Journal of Economic Psychology, 35 (2013), pp. 17–30.
- [50] M. CICERO, *The orations of Marcus Tullius Cicero*, vol. 3, London : George Bell & Sons, 1851. C.D. Younge trans.
- [51] M. S. CLARK AND J. R. MILLS, *A theory of communal (and exchange) relationships*, Handbook of theories of social psychology, 2 (2011), pp. 232–250.
- [52] R. CROSON AND J. Y. SHANG, *The impact of downward social information on contribution decisions*, Experimental Economics, 11 (2008), pp. 221–233.
- [53] R. T. CROSON, *Theories of commitment, altruism and reciprocity : Evidence from linear public goods games*, Economic Inquiry, 45 (2007), pp. 199–216.
- [54] A. R. DAMASIO, *Descartes' error : Emotion, reason, and the human brain*, New York : Putnam, 1994.
- [55] J. DANA, D. M. CAIN, AND R. M. DAWES, *What you don't know won't hurt me : Costly (but quiet) exit in dictator games*, Organizational Behavior and human decision Processes, 100 (2006), pp. 193–201.

- [56] C. DARWIN, P. EKMAN, AND P. PRODGER, *The expression of the emotions in man and animals*, Oxford University Press, 1998.
- [57] R. M. DAWES, *Social dilemmas*, Annual review of psychology, 31 (1980), pp. 169–193.
- [58] I. E. DE HOOGHE, S. M. BREUGELMANS, AND M. ZEELENBERG, *Not so ugly after all : when shame acts as a commitment device.*, Journal of personality and social psychology, 95 (2008), pp. 933–943.
- [59] I. E. DE HOOGHE, R. NELISSEN, S. M. BREUGELMANS, AND M. ZEELENBERG, *What is moral about guilt ? acting "prosocially" at the disadvantage of others.*, Journal of personality and social psychology, 100 (2011), pp. 462–473.
- [60] I. E. DE HOOGHE, M. ZEELENBERG, AND S. M. BREUGELMANS, *Moral sentiments and cooperation : Differential influences of shame and guilt*, Cognition and emotion, 21 (2007), pp. 1025–1042.
- [61] E. L. DECI, H. EGHARI, B. C. PATRICK, AND D. R. LEONE, *Facilitating internalization : The self-determination theory perspective*, Journal of personality, 62 (1994), pp. 119–142.
- [62] E. L. DECI AND R. M. RYAN, *A motivational approach to self : Integration in personality*, vol. 38, Lincoln : University of Nebraska Press, 1991, pp. 237–288.
- [63] R. DESSI AND X. ZHAO, *Self-esteem, shame and personal motivation*, CEPR Discussion Paper No. DP8722, (2011).
- [64] D. DESTENO, M. Y. BARTLETT, J. BAUMANN, L. A. WILLIAMS, AND L. DICKENS, *Gratitude as moral sentiment : emotion-guided cooperation in economic exchange*, Emotion, 10 (2010), pp. 289–293.
- [65] D. DESTENO, Y. LI, L. DICKENS, AND J. S. LERNER, *Gratitude a tool for reducing economic impatience*, Psychological science, 25 (2014), pp. 1262–1267.
- [66] M. DUFWENBERG AND U. GNEEZY, *Measuring beliefs in an experimental lost wallet game*, Games and economic Behavior, 30 (2000), pp. 163–182.
- [67] J. R. DUNN AND M. E. SCHWEITZER, *Feeling and believing : the influence of emotion on trust*, Journal of personality and social psychology, 88 (2005), pp. 736–748.
- [68] J. E. DUTTON, J. M. DUKERICH, AND C. V. HARQUAIL, *Organizational images and member identification*, Administrative science quarterly, 39 (1994), pp. 239–263.
- [69] N. EBER AND M. WILLINGER, *L'économie expérimentale*, Repères, Paris : La découverte, nouvelle édition ed., 2012.
- [70] N. EISENBERG, *Altruistic Emotion, Cognition, and Behavior (PLE : Emotion)*, Psychology Press, 2014.
- [71] P. EKMAN AND K. R. SCHERER, *Studies in Emotion and Social Interaction*, Cambridge University Press, 1982.
- [72] J. ELSTER, *Emotions and economic theory*, Journal of economic literature, 36 (1998), pp. 47–74.
- [73] R. A. EMMONS AND C. A. CRUMPLER, *Gratitude as a human strength : Appraising the evidence*, Journal of Social and Clinical Psychology, 19 (2000), pp. 56–69.
- [74] R. A. EMMONS, M. E. MCCULLOUGH, AND J.-A. TSANG, *The assessment of gratitude*, Washington, DC, US : American Psychological Association, 2003, pp. 327–341.

- [75] R. A. EMMONS AND C. M. SHELTON, *Gratitude and the science of positive psychology*, Handbook of positive psychology, 18 (2002), pp. 459–471.
- [76] M. ERLEI, *Heterogeneous social preferences*, Journal of Economic Behavior & Organization, 65 (2008), pp. 436–457.
- [77] A. FALK, E. FEHR, AND U. FISCHBACHER, *On the nature of fair behavior*, Economic Inquiry, 41 (2003), pp. 20–26.
- [78] A. FALK AND U. FISCHBACHER, *A theory of reciprocity*, Games and economic behavior, 54 (2006), pp. 293–315.
- [79] E. FEHR AND U. FISCHBACHER, *Third-party punishment and social norms*, Evolution and human behavior, 25 (2004a), pp. 63–87.
- [80] —, *Social norms and human cooperation*, Trends in cognitive sciences, 8 (2004b), pp. 185–190.
- [81] E. FEHR AND S. GÄCHTER, *Cooperation and punishment in public goods experiments*, The American Economic Review, 90 (2000), pp. 980–994.
- [82] E. FEHR AND K. M. SCHMIDT, *A theory of fairness, competition, and cooperation*, Quarterly journal of Economics, 114 (1999), pp. 817–868.
- [83] —, *The economics of fairness, reciprocity and altruism-experimental evidence and new theories*, vol. 1, Elsevier, 2006, ch. 8, pp. 615–691.
- [84] S. FERRARI, M. M. MEKNI, E. PETIT, AND S. ROUILLON, *Du bien-fondé de la participation des citoyens aux marchés de permis d'émissions : efficacité économique et questionnements éthiques*, Vertigo-la revue électronique en sciences de l'environnement, 10 (2010).
- [85] N. D. FESHBACH, *Empathy in children : Some theoretical and empirical considerations*, The Counseling Psychologist, 5 (1975), pp. 25–30.
- [86] U. FISCHBACHER AND S. GÄCHTER, *Social preferences, beliefs, and the dynamics of free riding in public goods experiments*, The American economic review, 100 (2010), pp. 541–556.
- [87] U. FISCHBACHER, S. GÄCHTER, AND E. FEHR, *Are people conditionally cooperative ? evidence from a public goods experiment*, Economics letters, 71 (2001), pp. 397–404.
- [88] J. FLEMING, *Relating employee engagement and customer loyalty to business outcomes in the financial service area*, Gallup Research Journal, 3 (2000), pp. 91–101.
- [89] —, *Relating employee engagement and customer loyalty to business outcomes in the retail industry*, The Gallup Research Journal, 3 (2000), pp. 103–115.
- [90] C. L. FLINCHBAUGH, E. W. G. MOORE, Y. K. CHANG, AND D. R. MAY, *Student well-being interventions : The effects of stress management techniques and gratitude journaling in the management education classroom*, Journal of Management Education, 36 (2011), pp. 191–219.
- [91] J. H. FOWLER AND N. A. CHRISTAKIS, *Cooperative behavior cascades in human social networks*, Proceedings of the National Academy of Sciences, 107 (2010), pp. 5334–5338.
- [92] R. H. FRANK, *Passions within reason : the strategic role of the emotions*, WW Norton & Co, 1988.

- [93] B. L. FREDRICKSON, *The role of positive emotions in positive psychology : The broaden-and-build theory of positive emotions.*, American psychologist, 56 (2001), pp. 218–226.
- [94] —, *Gratitude, Like Other Positive Emotions, Broadens and Builds*, New York, NY, US : Oxford University Press, 2004, ch. 8, pp. 145–166.
- [95] B. L. FREDRICKSON, M. A. COHN, K. A. COFFEY, J. PEK, AND S. M. FINKEL, *Open hearts build lives : positive emotions, induced through loving-kindness meditation, build consequential personal resources*, Journal of personality and social psychology, 95 (2008), pp. 1045–1062.
- [96] B. S. FREY AND S. MEIER, *Social comparisons and pro-social behavior : Testing" conditional cooperation" in a field experiment*, The American Economic Review, 94 (2004), pp. 1717–1722.
- [97] M. B. FRISCH, *Improving mental and physical health care through quality of life therapy and assessment*, in Advances in quality of life theory and research, Springer, 2000, pp. 207–241.
- [98] J. J. FROH, W. J. SEFICK, AND R. A. EMMONS, *Counting blessings in early adolescents : An experimental study of gratitude and subjective well-being*, Journal of School Psychology, 46 (2008), pp. 213–233.
- [99] S. GÄCHTER, *Conditional Cooperation : Behavioral Regularities from the Lab and the Field and Their Policy Implications.*, Cambridge : The MIT Press, 2007, pp. 19–50.
- [100] S. GÄCHTER AND B. HERRMANN, *Reciprocity, culture and human cooperation : previous insights and a new cross-cultural experiment*, Philosophical Transactions of the Royal Society B : Biological Sciences, 364 (2009), pp. 791–806.
- [101] L. GAGNÉ, *La modélisation des comportements non conséquentialistes en théorie du choix rationnel*, Philosophiques, 34 (2007), pp. 329–352.
- [102] A. GARAPIN, *Rapprocher l'homo oeconomicus de l'homo sapiens. vers une théorie des jeux réaliste et prédictive des comportements humains*, Revue d'économie politique, 119 (2009), pp. 1–40.
- [103] J. GEANAKOPOLOS, D. PEARCE, AND E. STACCHETTI, *Psychological games and sequential rationality*, Games and economic Behavior, 1 (1989), pp. 60–79.
- [104] U. GNEEZY AND A. RUSTICHINI, *Fine is a price, a*, J. Legal Stud., 29 (2000), pp. 1–18.
- [105] N. J. GOLDSTEIN, R. B. CIALDINI, AND V. GRISKEVICIUS, *A room with a view-point : Using social norms to motivate environmental conservation in hotels*, Journal of consumer Research, 35 (2008), pp. 472–482.
- [106] A. M. GRANT AND F. GINO, *A little thanks goes a long way : Explaining why gratitude expressions motivate prosocial behavior.*, Journal of personality and social psychology, 98 (2010), pp. 946–955.
- [107] A. M. GRANT AND A. WRZESNIEWSKI, *I won't let you down... or will i ? core self-evaluations, other-orientation, anticipated guilt and gratitude, and job performance.*, Journal of Applied Psychology, 95 (2010), pp. 108–121.
- [108] M. S. GREENBERG AND D. R. WESTCOTT, *Indebtedness as a mediator of reactions to aid*, New directions in helping, 1 (1983), pp. 85–112.

- [109] J. GREENE AND J. HAIDT, *How (and where) does moral judgment work ?*, Trends in cognitive sciences, 6 (2002), pp. 517–523.
- [110] F. GUALA, *The methodology of experimental economics*, Cambridge University Press, 2005.
- [111] F. GUALA AND L. MITTONE, *Paradigmatic experiments : the dictator game*, The Journal of Socio-Economics, 39 (2010), pp. 578–584.
- [112] J. HAIDT, *The emotional dog and its rational tail : a social intuitionist approach to moral judgment.*, Psychological review, 108 (2001), pp. 814–834.
- [113] —, *The moral emotions*, Handbook of affective sciences, 11 (2003), pp. 852–870.
- [114] J. HARTEK, *The linkage of employee perception to outcomes in a retail environment : Cause and effect*, The Gallup Research Journal, 3 (2000), pp. 25–38.
- [115] J. K. HARTEK, F. L. SCHMIDT, AND T. L. HAYES, *Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes : a meta-analysis*, Journal of applied psychology, 87 (2002), pp. 268–279.
- [116] M. HASSEBRAUCK AND B. FEHR, *Dimensions of relationship quality*, Personal relationships, 9 (2002), pp. 253–270.
- [117] A. HAUSMAN, W. J. JOHNSTON, AND L. YOUNG, *Trust : looking forward and back*, Journal of Business & Industrial Marketing, 21 (2006), pp. 439–445.
- [118] M. J. HEISEL AND M. MONGRAIN, *Facial expressions and ambivalence : Looking for conflict in all the right faces*, Journal of Nonverbal Behavior, 28 (2004), pp. 35–52.
- [119] G. HERTEL, J. NEUHOF, T. THEUER, AND N. L. KERR, *Mood effects on cooperation in small groups : Does positive mood simply lead to more cooperation ?*, Cognition & emotion, 14 (2000), pp. 441–472.
- [120] E. HOFFMAN, K. A. MCCABE, AND V. L. SMITH, *On expectations and the monetary stakes in ultimatum games*, International Journal of Game Theory, 25 (1996), pp. 289–301.
- [121] H. HOLLÄNDER, *A social exchange approach to voluntary cooperation*, The American Economic Review, 80 (1990), pp. 1157–1167.
- [122] A. HOPFENSITZ AND E. REUBEN, *The importance of emotions for the effectiveness of social punishment*, The Economic Journal, 119 (2009), pp. 1534–1559.
- [123] D. HUME, *A Treatise of Human Nature*, Oxford : Clarendon Press, 1896. 3 vols [1740].
- [124] T. L. HUSTON AND G. LEVINGER, *Interpersonal attraction and relationships*, Annual review of psychology, 29 (1978), pp. 115–156.
- [125] A. M. ISEN AND K. A. DAUBMAN, *The influence of affect on categorization*, Journal of personality and social psychology, 47 (1984), pp. 1206–1217.
- [126] L. A. JACKSON, D. A. LEWANDOWSKI, R. E. FLEURY, AND P. P. CHIN, *Effects of affect, stereotype consistency, and valence of behavior on causal attributions*, The Journal of social psychology, 141 (2001), pp. 31–48.
- [127] M. JOFFILY, D. MASCLÉ, C. N. NOUSSAIR, AND M. C. VILLEVAL, *Emotions, sanctions, and cooperation*, Southern Economic Journal, 80 (2014), pp. 1002–1027.

- [128] G. R. JONES AND J. M. GEORGE, *The experience and evolution of trust : Implications for cooperation and teamwork*, *Academy of management review*, 23 (1998), pp. 531–546.
- [129] R. JOURDHEUIL AND E. PETIT, *Émotions morales et comportement prosocial : Une revue de la littérature*, *Revue d'économie politique*, 125 (2015), pp. 499–525.
- [130] T. JUDGE, E. LOCKE, AND C. DURHAM, *The dispositional causes of job satisfaction : A core evaluation approach*, *Research in organizational Behavior*, 19 (1997), pp. 151–188.
- [131] W. A. KAHN, *Psychological conditions of personal engagement and disengagement at work*, *Academy of management journal*, 33 (1990), pp. 692–724.
- [132] G. KÉDIA, *La morale et les émotions*, *Revue électronique de psychologie sociale*, 4 (2009), pp. 47–53.
- [133] D. KELTNER AND J. J. GROSS, *Functional accounts of emotions*, *Cognition & Emotion*, 13 (1999), pp. 467–480.
- [134] C. KESER AND F. VAN WINDEN, *Conditional cooperation and voluntary contributions to public goods*, *The Scandinavian Journal of Economics*, 102 (2000), pp. 23–39.
- [135] T. KETELAAR, *The role of moral sentiments in economic decision making*, *Social psychology and economics*, (2006), pp. 97–116.
- [136] T. KETELAAR AND W. TUNG AU, *The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games : An affect-as-information interpretation of the role of emotion in social interaction*, *Cognition & Emotion*, 17 (2003), pp. 429–453.
- [137] L. A. KING, *Ambivalence over emotional expression and reading emotions in situations and faces*, *Journal of Personality and Social Psychology*, 74 (1998), pp. 753–762.
- [138] I. KOLSTAD, *The evolution of social norms : With managerial implications*, *The Journal of Socio-Economics*, 36 (2007), pp. 58–72.
- [139] A. E. KOMTER, *Gratitude and Gift Exchange*, New York : Oxford University Press, 2004, ch. 10, pp. 195–212.
- [140] N. KRAUSE, L. A. EVANS, G. POWERS, AND R. D. HAYWARD, *Feeling grateful to god : A qualitative inquiry*, *The Journal of Positive Psychology*, 7 (2012), pp. 119–130.
- [141] N. M. LAMBERT, M. S. CLARK, J. DURTSCHI, F. D. FINCHAM, AND S. M. GRAHAM, *Benefits of expressing gratitude expressing gratitude to a partner changes one's view of the relationship*, *Psychological Science*, 21 (2010), pp. 574–580.
- [142] N. M. LAMBERT, F. D. FINCHAM, AND T. F. STILLMAN, *Gratitude and depressive symptoms : The role of positive reframing and positive emotion*, *Cognition & emotion*, 26 (2012), pp. 615–633.
- [143] N. M. LAMBERT, S. M. GRAHAM, F. D. FINCHAM, AND T. F. STILLMAN, *A changed perspective : How gratitude can affect sense of coherence through positive reframing*, *The Journal of Positive Psychology*, 4 (2009), pp. 461–470.
- [144] R. LAZARUS AND B. LAZARUS, *Passion and Reason. Making Sense of Our Emotions*, New York : Oxford University Press, 1994.

- [145] S. E. LEA AND P. WEBLEY, *Pride in economic psychology*, Journal of Economic Psychology, 18 (1997), pp. 323–340.
- [146] J. LEDYARD, *Public goods : A survey of experimental research*, Princeton : Princeton university press, 1995, ch. 2, pp. 111–181.
- [147] J. S. LERNER, Y. LI, AND E. U. WEBER, *The financial costs of sadness*, Psychological science, 24 (2012), pp. 72–79.
- [148] D. K. LEVINE, *Modeling altruism and spitefulness in experiments*, Review of economic dynamics, 1 (1998), pp. 593–622.
- [149] J. LI, E. XIAO, D. HOUSER, AND P. R. MONTAGUE, *Neural responses to sanction threats in two-party economic exchange*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 16835–16840.
- [150] W. B. LIEBRAND, *The effect of social motives, communication and group size on behaviour in an n-person multi-stage mixed-motive game*, European Journal of Social Psychology, 14 (1984), pp. 239–264.
- [151] R. LÓPEZ-PÉREZ, *Aversion to norm-breaking : A model*, Games and Economic behavior, 64 (2008), pp. 237–267.
- [152] L. M., *Shame : The Exposed Self*, New York : Free Press, 1992.
- [153] M. MANNER AND J. GOWDY, *The evolution of social and moral behavior : Evolutionary insights for public policy*, Ecological Economics, 69 (2010), pp. 753–761.
- [154] J. G. MARCH AND J. P. OLSEN, *Democratic governance*, Free Press, 1995.
- [155] D. MASCLET, C. NOUSSAIR, S. TUCKER, AND M.-C. VILLEVAL, *Monetary and nonmonetary punishment in the voluntary contributions mechanism*, The American Economic Review, 93 (2003), pp. 366–380.
- [156] D. MASCLET, C. N. NOUSSAIR, AND M.-C. VILLEVAL, *Threat and punishment in public good experiments*, Economic Inquiry, 51 (2013), pp. 1421–1441.
- [157] M. F. MASCOLO AND K. W. FISCHER, *Developmental transformations in appraisals for pride, shame, and guilt*, New York, NY, US : Guilford Press, 1995, pp. 64–113.
- [158] D. R. MAY, R. L. GILSON, AND L. M. HARTER, *The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work*, Journal of occupational and organizational psychology, 77 (2004), pp. 11–37.
- [159] D. J. MCALLISTER, *Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations*, Academy of management journal, 38 (1995), pp. 24–59.
- [160] M. E. MCCULLOUGH, *Forgiveness as human strength : Theory, measurement, and links to well-being*, Journal of Social and Clinical Psychology, 19 (2000), pp. 43–55.
- [161] M. E. MCCULLOUGH, C. G. BELLAH, S. D. KILPATRICK, AND J. L. JOHNSON, *Vengefulness : Relationships with forgiveness, rumination, well-being, and the big five*, Personality and Social Psychology Bulletin, 27 (2001), pp. 601–610.
- [162] M. E. MCCULLOUGH, R. A. EMMONS, AND J.-A. TSANG, *The grateful disposition : a conceptual and empirical topography*, Journal of personality and social psychology, 82 (2002), pp. 112–127.
- [163] M. E. MCCULLOUGH, S. D. KILPATRICK, R. A. EMMONS, AND D. B. LARSON, *Is gratitude a moral affect ?*, Psychological bulletin, 127 (2001), pp. 249–266.

- [164] M. E. McCULLOUGH, M. B. KIMELDORF, AND A. D. COHEN, *An adaptation for altruism the social causes, social effects, and social evolution of gratitude*, Current directions in psychological science, 17 (2008), pp. 281–285.
- [165] M. E. McCULLOUGH, J.-A. TSANG, AND R. A. EMMONS, *Gratitude in intermediate affective terrain : links of grateful moods to individual differences and daily emotional experience*, Journal of personality and social psychology, 86 (2004), pp. 295–309.
- [166] M. E. McCULLOUGH, E. L. WORTHINGTON JR, AND K. C. RACHAL, *Interpersonal forgiving in close relationships.*, Journal of personality and social psychology, 73 (1997), pp. 321–336.
- [167] M. McPHERSON, L. SMITH-LOVIN, AND J. M. COOK, *Birds of a feather : Homophily in social networks*, Annual review of sociology, (2001), pp. 415–444.
- [168] R. S. MILLER, *The nature and severity of self-reported embarrassing circumstances*, Personality and Social Psychology Bulletin, 18 (1992), pp. 190–198.
- [169] ———, *On the nature of embarrassability : Shyness, social evaluation, and social skill*, Journal of personality, 63 (1995), pp. 315–339.
- [170] W. I. MILLER, *The anatomy of disgust*, Harvard University Press, 1998.
- [171] J. MILLS, M. S. CLARK, T. E. FORD, AND M. JOHNSON, *Measurement of communal strength*, Personal Relationships, 11 (2004), pp. 213–230.
- [172] L. MORETTI AND G. DI PELLEGRINO, *Disgust selectively modulates reciprocal fairness in economic interactions.*, Emotion, 10 (2010), pp. 169–180.
- [173] M. W. MORRIS AND D. KELTNER, *How emotions work : The social functions of emotional expression in negotiations*, Research in organizational behavior, 22 (2000), pp. 1–50.
- [174] B. MUNIER AND C. ZAHARIA, *High stakes and acceptance behavior in ultimatum bargaining*, Theory and Decision, 53 (2002), pp. 187–207.
- [175] W. S. NEILSON, *A theory of kindness, reluctance, and shame for social preferences*, Games and Economic Behavior, 66 (2009), pp. 394–403.
- [176] F. N.H., *The Emotions (Studies in Emotion and Social Interaction)*, New York : Cambridge University Press, 1986.
- [177] P. M. NIEDENTHAL, S. KRAUTH-GRUBER, AND F. RIC, *Psychology of emotion : Interpersonal, experiential, and cognitive approaches.*, Psychology Press, 2006.
- [178] M. A. NOWAK AND S. ROCH, *Upstream reciprocity and the evolution of gratitude*, Proceedings of the Royal Society of London B : Biological Sciences, 274 (2007), pp. 605–610.
- [179] T. OFFERMAN, J. SONNEMANS, AND A. SCHRAM, *Value orientations, expectations and voluntary contributions in public goods*, The economic journal, 106 (1996), pp. 817–845.
- [180] A. ORLÉAN, *Jeux évolutionnistes et normes sociales*, Économie appliquée, 50 (1997), pp. 177–198.
- [181] A. ORTONY, G. L. CLORE, AND A. COLLINS, *The cognitive structure of emotions*, Cambridge university press, 1990.
- [182] R. W. PALMATIER, C. B. JARVIS, J. R. BECHKOFF, AND F. R. KARDES, *The role of customer gratitude in relationship marketing*, Journal of marketing, 73 (2009), pp. 1–18.

- [183] N. PARK, C. PETERSON, AND M. E. SELIGMAN, *Strengths of character and well-being*, Journal of social and Clinical Psychology, 23 (2004), pp. 603–619.
- [184] R. PEETERS AND M. VORSATZ, *Immaterial rewards and sanctions in a voluntary contribution experiment*, Economic Inquiry, 51 (2013), pp. 1442–1456.
- [185] E. PETIT, *Le rôle des affects en économie*, Revue d'économie politique, 119 (2009), pp. 859–897.
- [186] —, *La négociation : les enseignements du jeu de l'ultimatum*, Négociations, (2010), pp. 77–96.
- [187] —, *Economie des émotions*, la Découverte, 2015.
- [188] E. PETIT AND S. ROUILLON, *La négociation : les enseignements du jeu du dictateur*, Négociations, (2010), pp. 71–95.
- [189] J. PIAGET, *The moral judgement of the child*, Presses Universitaires de France, 1973.
- [190] D. PIETRONI, G. A. VAN KLEEF, C. K. DE DREU, AND S. PAGLIARO, *Emotions as strategic information : Effects of other's emotional expressions on fixed-pie perception, demands, and integrative behavior in negotiation*, Journal of Experimental Social Psychology, 44 (2008), pp. 1444–1454.
- [191] M. RABIN, *Incorporating fairness into game theory and economics*, The American economic review, 83 (1993), pp. 1281–1302.
- [192] D. G. RAND, A. DREBER, T. ELLINGSEN, D. FUDENBERG, AND M. A. NOWAK, *Positive interactions promote public cooperation*, Science, 325 (2009), pp. 1272–1275.
- [193] A. F. REESON AND J. G. TISELL, *Institutions, motivations and public goods : An experimental test of motivational crowding*, Journal of Economic Behavior & Organization, 68 (2008), pp. 273–281.
- [194] E. REUBEN AND A. RIEDL, *Enforcement of contribution norms in public good games with heterogeneous populations*, Games and Economic Behavior, 77 (2013), pp. 122–137.
- [195] E. REUBEN AND F. VAN WINDEN, *Fairness perceptions and prosocial emotions in the power to take*, Journal of Economic Psychology, 31 (2010), pp. 908–922.
- [196] J. K. RILLING AND A. G. SANFEY, *The neuroscience of social decision-making*, Annual review of psychology, 62 (2011), pp. 23–48.
- [197] J. H. ROZIN P. AND C. MCCAULEY, *Disgust*, New York : Guilford Press, 2000, ch. 40.
- [198] R. M. RYAN AND E. L. DECI, *Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.*, American psychologist, 55 (2000), pp. 68–78.
- [199] A. G. SANFEY, J. K. RILLING, J. A. ARONSON, L. E. NYSTROM, AND J. D. COHEN, *The neural basis of economic decision-making in the ultimatum game*, Science, 300 (2003), pp. 1755–1758.
- [200] T. J. SCHEFF, *Shame and conformity : The deference-emotion system*, American sociological review, (1988), pp. 395–406.
- [201] T. SCHELLING, *The Strategy of Conflict*, Cambridge, MA : Harvard University Press, 1960.

- [202] S. SCHNALL, J. ROPER, AND D. M. FESSLER, *Elevation leads to altruistic behavior*, *Psychological science*, 21 (2010), pp. 315–320.
- [203] M. SEFTON, R. SHUPP, AND J. M. WALKER, *The effect of rewards and sanctions in provision of public goods*, *Economic inquiry*, 45 (2007), pp. 671–690.
- [204] A. SEN, *Rationality and Freedom*, Cambridge, MA : Belknap Press, 2002.
- [205] S. SEN AND S. AIRIAU, *Emergence of norms through social learning.*, in *IJCAI*, vol. 7, 2007, pp. 1507–1512.
- [206] D. SERRA, *Sentiments moraux et économie expérimentale*, *Leçons de philosophie économique*, 3 (2007), pp. 543–590.
- [207] R. SETHI AND E. SOMANATHAN, *Preference evolution and reciprocity*, *Journal of economic theory*, 97 (2001), pp. 273–297.
- [208] K. M. SHELDON, R. M. RYAN, L. J. RAWSTHORNE, AND B. ILARDI, *Trait self and true self : Cross-role variation in the big-five personality traits and its relations with psychological authenticity and subjective well-being.*, *Journal of personality and social psychology*, 73 (1997), pp. 1380–1393.
- [209] G. SIMMEL AND K. H. WOLFF, *The sociology of georg simmel*, vol. 92892, Simon and Schuster, 1950.
- [210] R. SLONIM AND A. E. ROTH, *Learning in high stakes ultimatum games : An experiment in the slovak republic*, *Econometrica*, 66 (1998), pp. 569–596.
- [211] A. SMITH, *The Theory of Moral Sentiments*, London : A. Millar, 6th edn. ed., 1759. Available from the Library of Economics and Liberty.
- [212] A. SMITH, *Recherches sur la nature et les causes de la richesse des nations*, Paris : Flammarion, 1776.
- [213] V. L. SMITH, *An experimental study of competitive market behavior*, *The Journal of Political Economy*, 70 (1962), pp. 111–137.
- [214] J. SONNEMANS, F. VAN DIJK, AND F. VAN WINDEN, *On the dynamics of social ties structures in groups*, *Journal of economic psychology*, 27 (2006), pp. 187–204.
- [215] C. S. SRIPADA AND S. STICH, *A framework for the psychology of moral norms*, Oxford : Oxford University Press, 2006, ch. 17.
- [216] M. STEIN, *Gratitude and attitude : A note on emotional welfare*, *Social Psychology Quarterly*, 52 (1989), pp. 242–248.
- [217] N. STROHMINGER, R. L. LEWIS, AND D. E. MEYER, *Divergent effects of different positive emotions on moral judgment*, *Cognition*, 119 (2011), pp. 295–300.
- [218] J. SULS, S. WITENBERG, AND D. GUTKIN, *Evaluating reciprocal and nonreciprocal prosocial behavior developmental changes*, *Personality and Social Psychology Bulletin*, 7 (1981), pp. 25–31.
- [219] M. SUTTER, *Outcomes versus intentions : On the nature of fair behavior and its development with age*, *Journal of Economic Psychology*, 28 (2007), pp. 69–78.
- [220] M. SUTTER, S. HAIGNER, AND M. G. KOCHER, *Choosing the carrot or the stick ? endogenous institutional choice in social dilemma situations*, *The Review of Economic Studies*, 77 (2010), pp. 1540–1566.
- [221] A. SZOLNOKI AND M. PERC, *Reward and cooperation in the spatial public goods game*, *EPL (Europhysics Letters)*, 92 (2010), pp. 1–6.

- [222] S. TADELIS, *The power of shame and the rationality of trust*, Haas School of Business Working Paper, (2011).
- [223] J. TANGNEY, D. MARSCHALL, K. ROSENBERG, D. BARLOW, AND P. WAGNER, *Children's and adults' autobiographical accounts of shame, guilt, and pride experiences : A qualitative analysis of situational determinants and interpersonal concerns*, Unpublished manuscript, George Mason University, (1994).
- [224] J. P. TANGNEY, J. STUEWIG, AND D. J. MASHEK, *Moral emotions and moral behavior*, Annual review of psychology, 58 (2007), pp. 345–372.
- [225] R. L. TRIVERS, *The evolution of reciprocal altruism*, Quarterly review of biology, 46 (1971), pp. 35–57.
- [226] J.-A. TSANG, *Brief report gratitude and prosocial behaviour : An experimental test of gratitude*, Cognition & Emotion, 20 (2006a), pp. 138–148.
- [227] ———, *The effects of helper intention on gratitude and indebtedness*, Motivation and Emotion, 30 (2006b), pp. 198–204.
- [228] A. W. TUCKER, *A two-person dilemma*, Mimeo, Stanford University, 1950.
- [229] T. VACHARKULKSEMSUK, L. E. SEKERKA, AND B. L. FREDRICKSON, *Establishing a positive emotional climate to create 21st-century organizational change*, The handbook of organizational culture and climate, 2 (2011), pp. 101–118.
- [230] P. VALDESOLO AND D. DESTENO, *Manipulations of emotional context shape moral judgment*, Psychological science, 17 (2006), pp. 476–477.
- [231] F. VAN DIJK, J. SONNEMANS, AND F. VAN WINDEN, *Social ties in a public good experiment*, Journal of Public Economics, 85 (2002), pp. 275–299.
- [232] F. VAN DIJK AND F. VAN WINDEN, *Dynamics of social ties and local public good provision*, Journal of Public Economics, 64 (1997), pp. 323–341.
- [233] D. VAN HOOREBEKE, *L'émotion et la prise de décision*, Revue française de gestion, (2008), pp. 33–44.
- [234] P. A. VAN LANGE, J. JOIREMAN, C. D. PARKS, AND E. VAN DIJK, *The psychology of social dilemmas : A review*, Organizational Behavior and Human Decision Processes, 120 (2013), pp. 125–141.
- [235] F. VAN WINDEN, M. STALLEN, AND K. R. RIDDERINKHOF, *On the nature, modeling, and neural bases of social ties*, vol. 20, Emerald Group Publishing, 2008.
- [236] C. VANBERG, *Why do people keep their promises? an experimental test of two explanations¹*, Econometrica, 76 (2008), pp. 1467–1480.
- [237] V. VANBERG ET AL., *Rationality, Rule Following and Emotions : On the Economics of Moral Preferences*, Max Planck Inst. of Economics, Evolutionary Economics Group, 2006.
- [238] V. J. VANBERG, *On the economics of moral preferences*, American journal of Economics and Sociology, 67 (2008), pp. 605–628.
- [239] R. VEENHOVEN, *How harmful is happiness*, Rotterdam : Universitaire Pers Rotterdam, (1989).
- [240] A. E. VÉLEZ GARCÍA AND F. OSTROSKY-SOLÍS, *From morality to moral emotions*, International journal of psychology, 41 (2006), pp. 348–354.
- [241] M. C. VILLEVAL, *Quand le marché ne suffit plus : biens publics et coopération conditionnelle*, Idées économiques et sociales, (2010), pp. 6–14.

- [242] K. D. VOHS, R. F. BAUMEISTER, AND J. CHIN, *Feeling duped : Emotional, motivational, and cognitive aspects of being exploited by others.*, Review of General Psychology, 11 (2007), pp. 127–141.
- [243] J. M. WALKER AND M. A. HALLORAN, *Rewards and sanctions and the provision of public goods in one-shot settings*, Experimental Economics, 7 (2004), pp. 235–247.
- [244] C. J. WATKINS AND P. DAYAN, *Q-learning*, Machine learning, 8 (1992), pp. 279–292.
- [245] P. WATKINS, J. SCHEER, M. OVNICEK, AND R. KOLTS, *The debt of gratitude : Dissociating gratitude and indebtedness*, Cognition & Emotion, 20 (2006), pp. 217–241.
- [246] P. C. WATKINS, *Gratitude and Subjective Well-Being*, New York : Oxford University Press, 2004, ch. 9, pp. 167–192.
- [247] C. E. WAUGH AND B. L. FREDRICKSON, *Nice to know you : Positive emotions, self–other overlap, and complex understanding in the formation of a new relationship*, The Journal of Positive Psychology, 1 (2006), pp. 93–106.
- [248] B. WOJCISZKE, *Morality and competence in person-and self-perception*, European review of social psychology, 16 (2005), pp. 155–188.
- [249] A. M. WOOD, G. D. BROWN, AND J. MALTBY, *Thanks, but i’m used to better : a relative rank model of gratitude.*, Emotion, 11 (2011), pp. 175–180.
- [250] A. M. WOOD, J. J. FROH, AND A. W. GERAGHTY, *Gratitude and well-being : A review and theoretical integration*, Clinical psychology review, 30 (2010), pp. 890–905.
- [251] A. M. WOOD, S. JOSEPH, AND J. MALTBY, *Gratitude uniquely predicts satisfaction with life : Incremental validity above the domains and facets of the five factor model*, Personality and Individual Differences, 45 (2008), pp. 49–54.
- [252] ———, *Gratitude predicts psychological well-being above the big five facets*, Personality and Individual Differences, 46 (2009), pp. 443–447.
- [253] A. M. WOOD, J. MALTBY, R. GILLETT, P. A. LINLEY, AND S. JOSEPH, *The role of gratitude in the development of social support, stress, and depression : Two longitudinal studies*, Journal of Research in Personality, 42 (2008), pp. 854–871.
- [254] T. A. WRIGHT AND R. CROPANZANO, *The role of psychological well-being in job performance : : A fresh look at an age-old quest*, Organizational Dynamics, 33 (2004), pp. 338–351.
- [255] E. XIAO AND D. HOUSER, *Emotion expression in human punishment behavior*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), pp. 7398–7401.
- [256] T. YAMAGISHI AND N. MIFUNE, *Does shared group membership promote altruism ? fear, greed, and reputation*, Rationality and Society, 20 (2008), pp. 5–30.
- [257] M. ZEELLENBERG AND R. PIETERS, *A theory of regret regulation 1.0*, Journal of Consumer psychology, 17 (2007), pp. 3–18.
- [258] R. ZULTAN, *Strategic and social pre-play communication in the ultimatum game*, Journal of Economic Psychology, 33 (2012), pp. 425–434.

Sixième partie .

Annexe

Démonstration de la proposition 3

On suppose $a = \frac{1}{2}$ et $\bar{y}_i > \bar{y}_j$.

Par ailleurs, on rappelle que, pour tout $i \in I$, on suppose que $\delta_i \leq 1$ et que $2\beta_i > a\alpha_i$ (cf. 5.2.1.1.3).

De plus, pour tout $i, j \in I^2$, $i \neq j$, l'équilibre de Nash dans le JBP standard à $n = 2$ joueurs est :

$$y_i^* = \frac{a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i]}{4\beta_i\beta_j - a^2\alpha_i\alpha_j}$$

Ainsi, $y_i^* < \bar{y}_i$ est équivalent à :

$$\frac{a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i]}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} < \bar{y}_i$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i] < \bar{y}_i(4\beta_i\beta_j - a^2\alpha_i\alpha_j)$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i)] + 2\beta_j(2\beta_i - a\alpha_i)\bar{y}_i < \bar{y}_i(4\beta_i\beta_j - a^2\alpha_i\alpha_j)$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i)] < \bar{y}_i(2\beta_i2\beta_j - a\alpha_i a\alpha_j) - 2\beta_j(2\beta_i - a\alpha_i)\bar{y}_i$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + a\alpha_i(2\beta_j - a\alpha_j)\bar{y}_j + 2\beta_j[-1 + a(1 + \delta_i)] < (2\beta_i2\beta_j - a\alpha_i a\alpha_j - 2\beta_j2\beta_i + 2\beta_j a\alpha_i)\bar{y}_i$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + a\alpha_i(2\beta_j - a\alpha_j)\bar{y}_j + 2\beta_j[-1 + a(1 + \delta_i)] < a\alpha_i(2\beta_j - a\alpha_j)\bar{y}_i$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] < a\alpha_i(2\beta_j - a\alpha_j)\bar{y}_i - a\alpha_i(2\beta_j - a\alpha_j)\bar{y}_j$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] < a\alpha_i(2\beta_j - a\alpha_j)(\bar{y}_i - \bar{y}_j)$$

Soit

$$\frac{a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)]}{a\alpha_i(2\beta_j - a\alpha_j)} < (\bar{y}_i - \bar{y}_j)$$

D'après nos hypothèses, on sait que, pour tout $i \in I$, $a(1 + \delta_i) \leq 1$

Par conséquent, $a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] \leq 0$.

Et puisque l'on suppose que $\bar{y}_i - \bar{y}_j > 0$, alors $y_i^* < \bar{y}_i$.

Ainsi, lorsque $a = \frac{1}{2}$, la contribution optimale du joueur ayant la norme de contribution la plus importante (i ici), est toujours inférieure à la valeur de cette norme (\bar{y}_i ici).

Supposons maintenant que $\bar{y}_i < \bar{y}_j$ (on se place toujours du point de vue du joueur i), $y_i^* < \bar{y}_j$ est équivalent à :

$$\frac{a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i]}{4\beta_i\beta_j - a^2\alpha_i\alpha_j} < \bar{y}_j$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j) + (2\beta_j - a\alpha_j)\bar{y}_j] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i] < \bar{y}_j(4\beta_i\beta_j - a^2\alpha_i\alpha_j)$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + a\alpha_i(2\beta_j - a\alpha_j)\bar{y}_j + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i] < \bar{y}_j(2\beta_i2\beta_j - a\alpha_i a\alpha_j)$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i) + (2\beta_i - a\alpha_i)\bar{y}_i] < (2\beta_i2\beta_j - a\alpha_i a\alpha_j - a\alpha_i(2\beta_j - a\alpha_j))\bar{y}_j$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] + 2\beta_j(2\beta_i - a\alpha_i)\bar{y}_i < (2\beta_i2\beta_j - a\alpha_i a\alpha_j - a\alpha_i2\beta_j + a\alpha_i a\alpha_j)\bar{y}_j$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] + 2\beta_j(2\beta_i - a\alpha_i)\bar{y}_i < 2\beta_j(2\beta_i - a\alpha_i)\bar{y}_j$$

$$\Leftrightarrow a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] < 2\beta_j(2\beta_i - a\alpha_i)(\bar{y}_j - \bar{y}_i)$$

Soit

$$\frac{a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)]}{2\beta_j(2\beta_i - a\alpha_i)} < \bar{y}_j - \bar{y}_i$$

D'après nos hypothèses, on sait que $a\alpha_i[-1 + a(1 + \delta_j)] + 2\beta_j[-1 + a(1 + \delta_i)] \leq 0$.

Et puisque l'on suppose que $\bar{y}_j - \bar{y}_i > 0$, alors $y_i^* < \bar{y}_i$.

Ainsi, lorsque $a = \frac{1}{2}$, la contribution optimale du joueur ayant la norme de contribution la moins importante (i ici), est toujours inférieure à la valeur de la norme de contribution de son partenaire (\bar{y}_j ici).

On montre ainsi que, dans le JBP standard à $n = 2$ joueurs, lorsque $a = \frac{1}{2}$, la contribution des agents, à l'équilibre, n'atteint jamais le niveau de la norme de contribution la plus élevée (entre les deux partenaires).