

# La modélisation des comportements non conséquentialistes en théorie du choix rationnel

Learry Gagné

Volume 34, Number 2, Fall 2007

URI: <https://id.erudit.org/iderudit/016995ar>

DOI: <https://doi.org/10.7202/016995ar>

[See table of contents](#)

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)

1492-1391 (digital)

[Explore this journal](#)

Cite this article

Gagné, L. (2007). La modélisation des comportements non conséquentialistes en théorie du choix rationnel. *Philosophiques*, 34(2), 329–352.  
<https://doi.org/10.7202/016995ar>

Article abstract

In this paper, we attempt to find the place of non-consequentialist behavior, like respect of norms and values, in rational-choice theory. At the onset, there is no limit as to what can constitute a preference or a utility value; any non-consequentialist behavior can be reduced to a consequentialist one. A brief survey of several rational models of social norms shows us, on one hand, that consequentialist reduction of conformity leaves some important phenomena unexplained, and, on the other hand, that recourse to non-instrumental utilities generates serious problems within the theory. We end with a discussion of two alternative approaches to rational-choice theory that leave a large place for non-consequentialist behavior. We conclude by saying that if, for the same action, the type of motivation behind it (consequentialist or not) has by itself an influence on the outcome of collective action, then in such cases we have to step out of rational-choice theory, without however abandoning it.

# La modélisation des comportements non conséquentialistes en théorie du choix rationnel<sup>1</sup>

LEARRY GAGNÉ

Université du Québec à Montréal

gagnel@awale.qc.ca

**RÉSUMÉ.** — Nous tentons, dans cet article, de déterminer la place des comportements non conséquentialistes, notamment le respect des valeurs et des normes, dans la théorie du choix rationnel. Au départ, il n'y a pas de limites à ce qui peut constituer une préférence ou une valeur d'utilité; tout comportement non conséquentialiste peut être réduit à un comportement conséquentialiste. Un bref examen de certains modèles rationnels des normes sociales nous montre, d'une part, que la réduction conséquentialiste du conformisme laisse inexplicables certains phénomènes importants, et d'autre part, que l'emploi d'utilités autres qu'instrumentales pose de sérieux problèmes à la théorie. Pour terminer, nous discutons de deux autres approches de la théorie du choix rationnel faisant une large part aux comportements non conséquentialistes. Nous concluons que si, pour une même action, le type de motivation (conséquentialiste ou non) influence en soi les résultats de l'action collective, alors il nous faut dans ce cas sortir du cadre de la rationalité, sans toutefois complètement l'abandonner.

**ABSTRACT.** — In this paper, we attempt to find the place of non-consequentialist behavior, like respect of norms and values, in rational-choice theory. At the onset, there is no limit as to what can constitute a preference or a utility value; any non-consequentialist behavior can be reduced to a consequentialist one. A brief survey of several rational models of social norms shows us, on one hand, that consequentialist reduction of conformity leaves some important phenomena unexplained, and, on the other hand, that recourse to non-instrumental utilities generates serious problems within the theory. We end with a discussion of two alternative approaches to rational-choice theory that leave a large place for non-consequentialist behavior. We conclude by saying that if, for the same action, the type of motivation behind it (consequentialist or not) has by itself an influence on the outcome of collective action, then in such cases we have to step out of rational-choice theory, without however abandoning it.

Les succès de la théorie du choix rationnel en sciences sociales sont indéniables. Nous constatons toutefois que les modèles rationnels s'avèrent plus convaincants dans des champs où la motivation des agents est manifestement conséquentialiste et maximisatrice (les transactions de marché par exemple) que dans des champs où ce genre de motivation est moins évident (les pratiques religieuses, le choix d'un époux, etc.). Certains adeptes de la théorie répliqueront que les motivations réelles ne les intéressent pas, car leur méthode consiste à faire

---

1. J'aimerais remercier Robert Nadeau et Dominique Leydet pour leur soutien et leurs précieux commentaires lors de la rédaction de cet article.

« comme si » les agents étaient rationnels. D'autres affirmeront que les motivations non conséquentialistes importent, il s'agit de les traiter comme utilités spéciales au sein de la théorie. Dans cet article, nous nous demandons si, effectivement, la théorie du choix rationnel peut traiter des comportements non conséquentialistes de manière satisfaisante.

La théorie du choix rationnel s'intéresse avant tout aux comportements conséquentialistes : je vise un certain but, et j'applique les meilleurs moyens permettant de l'atteindre, considérant les coûts et les bénéfices des différentes possibilités. Un comportement non conséquentialiste est de nature « je fais X parce qu'il le faut ». Prenons l'exemple de deux témoins d'un crime questionnés par un juge, dans une cause où dire la vérité pourrait s'avérer incriminant. Le témoin A n'exhibe pas de préférences *a priori* pour l'honnêteté mais choisit de dire la vérité, car il croit que les sanctions contre une fausse déclaration seraient plus coûteuses que les inconvénients de l'honnêteté<sup>2</sup>. Le témoin B est mû par une valeur d'honnêteté ; il choisit de dire la vérité « parce qu'il le faut », comme on dit. Ces deux choix peuvent être modélisés comme une maximisation d'utilité : A recherche la plus petite peine possible, et B cherche à maximiser l'honnêteté. Du point de vue de la rationalité, a-t-on raison d'aborder ces deux agents de la même manière ? On remarque au départ que le fait que l'agent maximise son utilité ne constitue pas en soi une preuve suffisante de sa rationalité ; il faut également qu'il choisisse les moyens adéquats aux fins recherchées<sup>3</sup>. Or nous constatons que A choisit la vérité comme moyen en vue d'une fin indépendante, l'évitement de sanctions, alors que B confond moyens et fins, car il choisit la vérité dans le but d'être honnête, c'est-à-dire quelqu'un qui dit la vérité. C'est ce dernier type de choix que nous qualifions de « non conséquentialiste », car même s'il est d'une certaine manière maximisant, il est effectué pour lui-même, et non en vertu d'un but qui lui est extérieur. Ne serait-il pas plus approprié de considérer l'honnêteté comme une motivation distincte des motivations instrumentales ? Au lieu de voir l'action honnête comme un moyen calculé en vue de satisfaire une fin recherchée, l'honnêteté, nous pourrions concevoir une telle action simplement comme une manière d'agir qui ne se veut pas nécessairement rationnelle.

Les comportements de nature non conséquentialiste qui nous intéressent avant tout sont ceux qui peuvent être aisément redécrits en termes de rationalité. Nous laisserons de côté les comportements à la fois non conséquentialistes

---

2. On pourrait également dire qu'il préfère l'honnêteté au mensonge *ceteris paribus*, tout en étant particulièrement sensible à la structure de coûts, de sorte qu'il puisse rapidement choisir de mentir dès que la vérité s'avère désavantageuse.

3. La théorie de l'utilité contient en elle-même le critère de maximisation : lorsque l'agent classe ses options sur une échelle d'utilité, ce n'est logiquement pas dans le but de choisir le second sur la liste. Dans le contexte de l'utilité formelle, la rationalité par la maximisation devient tautologique (Hausman, 1992, 18), d'où la nécessité de recourir à l'adéquation des moyens aux fins de la définition de la rationalité (Demeulenaere, 1996, 240).

et considérés comme irrationnels dans la conception classique de la rationalité<sup>4</sup>. Nous porterons notre attention sur les actions qui ont un but, lorsque ce but se décrit dans le langage de l'action, de sorte que l'action ne semble pas être entreprise pour une autre raison que sa propre réalisation. Le champ des valeurs et des normes nous semble exhiber cette propriété : lorsque l'agent est motivé par une valeur fortement intériorisée, il a tendance à agir au nom de cette valeur, sans égard aux bénéfices potentiels autres que le respect de cette valeur. Le second témoin de la petite histoire mentionnée plus haut dit la vérité, car l'honnêteté est pour lui une valeur ; les sanctions contre une fausse déclaration n'entrent pas dans son raisonnement. Nous nous pencherons sur les valeurs et leur traitement en théorie du choix rationnel, mais uniquement sur cet aspect non conséquentialiste, et non sur la distinction entre intérêts égoïstes et non égoïstes parfois soulevée qui, sans être inintéressante, nous éloignerait de notre propos. Dans ce qui suit, nous effectuerons un survol des types de modélisation rationnelle des comportements non conséquentialistes. Nous débiterons par un débat épistémologique sur la pertinence de distinguer ces comportements. Nous ferons par la suite l'examen d'une série de tentatives en choix rationnel de modélisation des valeurs qui leur confèrent un statut particulier par rapport aux préférences « ordinaires ». Pour terminer, nous approfondirons la discussion à l'aide de deux approches critiques de la théorie du choix rationnel, qui prennent au sérieux la distinction entre conséquentialisme et non conséquentialisme, soit les alternatives proposées par Raymond Boudon et Jon Elster.

### La portée de la théorie du choix rationnel

Il faut toujours garder à l'esprit que, dans la théorie classique du choix rationnel, il n'y a aucune limite à ce qui constitue une préférence ou une valeur d'utilité ; donc, en dernière instance, la question des comportements non conséquentialistes ne s'y pose pas, car on peut toujours faire « comme s'ils » étaient conséquentialistes, pour autant qu'ils ne soient pas manifestement irrationnels. Dès lors se pose toutefois la question de la portée empirique de la théorie. On peut l'aborder de deux manières. Prenant l'agent comme point d'ancrage, la critique empirique *en amont* cherche à déterminer si celui-ci est bel et bien rationnel, en d'autres termes s'il est réellement motivé par une norme de rationalité. On peut faire fi de cette critique en amont en adoptant une conception idéale de l'agent et en le supposant rationnel, même si l'on sait très bien que ce n'est pas toujours le cas. C'est l'*hypothèse de rationalité* : l'agent

---

4. En bref, nous postulons trois catégories de comportements irrationnels (ou non rationnels) : le comportement *forcé*, soit par la présence d'un seul élément de choix, d'une réaction impulsive (y compris les actions causées par les émotions), ou d'une autorité coercitive ; le comportement *habituel*, où l'agent ne change pas ses choix malgré un changement de situation ; et à l'inverse le comportement *aléatoire*, où l'agent change constamment d'idée malgré un environnement stable.

est supposé agir en tout temps en fonction de sa préférence la plus élevée<sup>5</sup>. La critique empirique se déplace alors *en aval* de l'agent ; on cherche à savoir si le modèle rationnel constitué de tels agents idéalisés nous permet effectivement de mieux comprendre certains phénomènes sociaux. La théorie du choix rationnel a tendance à prendre plus au sérieux ce genre de critique. Sur cette question, on peut classer ses variantes en deux catégories : les théories générales et spéciales (Goldthorpe, 2000 : 123-125). Avec les théories générales, on peut se permettre de modéliser à peu près n'importe quelle situation sociale comme une interaction d'agents rationnels ; on pense à « l'impérialisme économique » de l'École de Chicago, représentée par l'œuvre de Becker<sup>6</sup> et, plus récemment, par Levitt et Dubner (2005). Les théories spéciales, quant à elles, cherchent à modéliser le comportement des agents d'une manière autre que pleinement rationnelle lorsque cette hypothèse ne semble pas tenir la route ; nous croiserons certaines instances de théories spéciales dans la seconde partie. Pour ce qui est de la modélisation des comportements non conséquentialistes, on peut s'attendre à ce que la théorie du choix rationnel, du moins dans ses variantes spéciales, prenne en considération le réalisme des résultats du modèle, et non le réalisme du comportement attribué à l'agent.

La théorie du choix rationnel peut ramener tout comportement non conséquentialiste en comportement rationnel maximiseur en adoptant des motivations et des préférences *ad hoc*. À défaut de se poser la question du réalisme de l'entreprise, nous pouvons tenter de voir si l'hypothèse de rationalité tient bien le coup. L'hypothèse de rationalité est, bien entendu, un principe de charité. Dans l'explication du comportement, on suppose d'abord que l'agent maximise une utilité quelconque, et ce n'est qu'après avoir épuisé les ressources de la théorie du choix rationnel que nous sommes en mesure d'affirmer que le comportement s'avère forcé, habituel, ou aléatoire. Comme les préférences peuvent porter théoriquement sur n'importe quoi, les limites du principe de charité vont dépendre de la modélisation particulière. Si l'on observe que l'agent perd systématiquement son argent à la suite de ses choix, que l'on soit prêt à le déclarer irrationnel ou bien montrant une préférence pour la pauvreté dépend entièrement de ce que l'on veut modéliser.

Partons d'un exemple commun de comportement non conséquentialiste : suivre une tradition. Pour Goldthorpe (2000 : 128-129), le comportement traditionnel, adopter toujours la même stratégie face au même problème, peut s'avérer rationnel si cette stratégie s'avère efficace. Par contre, si la situation de l'agent change suffisamment pour que la stratégie traditionnelle cesse d'être

---

5. Cependant, certaines variantes de la théorie vont quand même chercher à justifier cette position épistémologique. Par exemple, on peut soutenir que la rationalité est une motivation réelle car, face à sa propre irrationalité, l'agent ressentira un malaise et tentera de se corriger (Follesdal, 1982, 316). Ce genre de postulat psychologique n'est cependant pas nécessaire, tant que l'hypothèse de la rationalité ne concerne que la cohérence des choix (Ferejohn, 2002 : 224).

6. Notamment pour Becker (1986, 111-112), les comportements non optimaux s'expliquent par des coûts « monétaires ou psychiques » qui ont pu échapper à l'observateur. De plus, l'agent n'est pas nécessairement conscient de sa propre motivation rationnelle.

efficace, et que celui-ci persiste dans ses choix, alors, si l'on veut demeurer fidèle à la théorie du choix rationnel, il nous faut postuler soit une irrationalité, soit une préférence pour la tradition, et cette dernière lui apparaît beaucoup trop *ad hoc* pour justifier une modélisation rationnelle. On retrouve un semblable questionnement dans le fameux « paradoxe du vote » : pourquoi se déplacer pour aller voter alors que les coûts sont non négligeables et que le bénéfice potentiel que notre vote fasse une différence au décompte final est si minime ? Brennan (1991 : 89 n. 14), prenant le camp de la rationalité, fait remarquer que l'on peut facilement expliquer le geste de voter à une élection en invoquant une préférence pour le vote. Une telle réduction des motivations, quoiqu'extrême, demeure légitime. À l'inverse, pour Etzioni (1988 : 43-45), la fusion fins-moyens que l'on retrouve typiquement dans les décisions morales constitue une raison suffisante pour les dissocier des décisions instrumentales. La coupure ne nous apparaît toutefois pas aussi nette. Même si l'agent motivé par une valeur forte ne porte pas *habituellement* attention aux coûts et aux bénéfices de l'action, il est aisé, pour n'importe quelle valeur, d'imaginer une situation où les coûts de l'action morale seront tellement élevés que l'on pourra prédire que l'agent intégrera ces coûts dans son raisonnement. Il adoptera alors une attitude *instrumentale* face au respect de la valeur, et sa motivation correspondra à ce que la théorie du choix rationnel avait toujours supposé<sup>7</sup>. Même la moralité « impérative » chez Etzioni ne peut pas être complètement dissociée de la rationalité instrumentale :

*People first sense an absolute command to act morally, but that does not mean that they will always heed it. That they are less likely to heed it if the costs are high does not indicate that there is no imperative ; indeed, all other things being equal, it is what drives up the costs. (Etzioni, 1986 : 167)<sup>8</sup>.*

L'hypothèse de rationalité étant un principe de charité, on peut le pousser aussi loin qu'on le désire. Ces théoriciens se distinguent avant tout par le degré de charité qu'ils sont prêts à accepter.

Certains admettent leur impuissance à prendre en défaut l'hypothèse de rationalité de l'agent sur une base conceptuelle. Pour Goldthorpe, lorsque le respect de la valeur domine inconditionnellement le choix que recommanderait une analyse en termes de coûts et de bénéfices, « *it would again seem best to acknowledge that the limits of applicability of RAT [Rational Action Theory] are reached* » (Goldthorpe, 2000 : 129). Il reconnaît cependant que le théoricien du choix rationnel pourrait « sauver » un tel comportement comme rationnel. Le choix des mots est important (« *it would seem best...* »), car par là,

---

7. Pettit (1995) nous fournit une image de la rationalité comme étant des « balises virtuelles » qui viennent limiter nos actions morales lorsqu'elles s'éloignent trop de notre intérêt personnel.

8. La dernière partie de la citation porte à confusion, car elle semble indiquer que le caractère impératif de la valeur augmente son propre coût d'exécution. La proposition a plus de sens si on lit « *...it is what drives up the cost threshold* ». L'impératif rend l'agent plus tolérant aux coûts qu'il endosse en agissant de façon morale.

Goldthorpe admet qu'il n'y a pas de critère objectif de rejet d'un modèle rationnel ; cela devient alors une question de jugement sur ce qui semble fonctionner le mieux comme modèle. D'autres abordent ce problème sensiblement de la même manière. Isaac (1997 : 560-63) définit les « règles fortes » comme des règles de conduite indépendantes des aspects de la situation de choix. Toutefois, les règles fortes ne constituent pas selon lui un problème « logique » pour la théorie du choix rationnel, car on peut toujours interpréter le choix en termes de maximisation, mais la « maximisation abstraite » qu'une telle approche suppose affaiblit l'aspect empirique de la théorie, car elle s'intéresse peu aux motivations. Dans la même veine, Hausman et McPherson affirment : « *Even though it may be possible formally to model the committed agent as maximizing utility, it seems enlightening not to do so* » (Hausman, McPherson, 1993 : 688), à cause du rôle de la moralité dans les motivations de l'agent. Ce sont là des jugements de valeur sur la pertinence de l'hypothèse de rationalité selon les circonstances, et non des critiques conceptuelles. On peut se poser de sérieuses questions cependant sur le caractère tautologique des explications rationnelles des comportements non conséquentialistes, comme expliquer le conformisme par une préférence pour le conformisme (Demeulenaere, 1996 : 247-269).

Il semble bien que si l'on veut demeurer fidèle à la théorie du choix rationnel, il y a toujours moyen de modéliser les valeurs comme des préférences ou des contraintes, et de les inclure dans le calcul d'utilité. Pour Brennan, tant que nous considérons les valeurs comme *données*, et il n'y a aucune raison de faire autrement en choix rationnel, elles ne posent aucun problème de modélisation. Ce n'est, selon lui, que lorsque les valeurs sont choisies, et choisies pour un but autre que la maximisation de son bien-être, qu'elles peuvent devenir problématiques. Ce serait là de toute façon un questionnement sur la formation des valeurs, et le choix rationnel ne s'en préoccupe pas. Nous ne pouvons qu'être d'accord avec sa conclusion : « *An agent's preferences are simply taken to be what they are revealed to be ; a more complex conception is, by the 'norms' of economic methodology, both empirically unnecessary and theoretically illogical* » (Brennan, 1991 : 93). La loi de l'offre et de la demande suppose que les agents soient motivés à payer, ou à obtenir le meilleur prix possible pour leurs biens, même si dans le monde réel les agents peuvent montrer d'autres motivations. On peut exiger que les théoriciens s'intéressent davantage aux motivations, mais il n'y a rien dans la théorie qui les y oblige. La théorie du choix rationnel se distingue formellement de la sociologie interprétative, mais on peut espérer que ses partisans « interprètent », de temps à autre, leurs sujets afin d'ajouter du réalisme à leurs modèles<sup>9</sup> (Ferejohn, 2002 : 227-28).

---

9. Satz et Ferejohn (1994) adoptent la position de l'« externalisme modéré » : la théorie n'a pas à s'intéresser aux motivations internes des agents, mais lorsque les motivations modélisées correspondent aux motivations réelles, la théorie ne peut que s'en porter mieux.

### Modèles rationnels des valeurs

Nous allons maintenant nous intéresser plus en détail aux efforts entrepris dans les modèles rationnels visant à traiter des valeurs et des normes. Nous regroupons ces modèles selon qu'ils traitent des valeurs comme des utilités standard, c'est-à-dire conceptuellement conformes en tous points aux utilités matérielles, ou comme des utilités non standard possédant des caractéristiques particulières les distinguant des utilités matérielles.

#### *Modèles rationnels standard*

Supposons que nous défendions la théorie du choix rationnel contre les critiques en amont ; nous acceptons donc que les agents soient modélisés d'une façon idéalement rationnelle. La critique se déplace alors en aval : les modèles rationnels d'action collective sont-ils toujours adéquats ? L'approche évidente serait d'évaluer le réalisme des résultats du modèle. Cependant, la théorie du choix rationnel ne se veut pas absolument réaliste, et même si le modèle dévie de l'empirisme, ses conclusions peuvent conserver un intérêt non négligeable. Le type de critique que nous offrons ici est différent, et nous croyons qu'il constitue une charge plus sérieuse contre certains modèles rationnels, particulièrement les modèles intégrant rationalité et normes. Cette critique va comme suit : si on suppose que les agents adoptent *consciemment* et *ouvertement* la motivation rationnelle que le modèle leur attribue, est-ce que celui-ci tient toujours ? La question devient pertinente lorsque, dans une dynamique d'action collective, pour une *même* action, l'agent reçoit un « paiement » différent selon que sa motivation se présentait de façon conséquentialiste ou non conséquentialiste. Dans un tel cas, un modèle rationnel qui convertit les motivations non conséquentialistes verra lui échapper des aspects importants de la dynamique. Nous soutenons que les actions collectives issues des normes sociales présentent cette caractéristique ; il y a une distinction à faire entre le conformisme sincère (non conséquentialiste) et instrumental aux normes, notamment dans la réception de telles motivations par autrui.

Nous adopterons une définition des normes assez commune en rationalité, celle de Pettit (1990, 731) : une régularité de comportement R est une norme lorsque presque tout le monde approuve que chacun adopte R (et désapprouve le contraire), et que cette approbation a un effet sur l'adoption de R. Il existe des normes auxquelles il est individuellement et immédiatement rationnel de se conformer ; respecter les feux de circulation, par exemple, est rationnel, peu importe ce que les autres font. Un modèle rationnel serait ici parfaitement adéquat, par contre on préfère souvent dans de telles situations parler d'une « convention » plutôt que d'une « norme ». Les normes dans les modèles rationnels servent le plus souvent à résoudre des problèmes d'action collective pour lesquels il n'est pas immédiatement rationnel de coopérer, mais lors d'un seuil critique où les agents coopèrent, un surplus bénéfique à tous est généré. Ce sont les cas classiques de production de biens publics et de Dilemmes du prisonnier. La coopération rationnelle peut être assurée par des



institutions établissant un système de sanctions<sup>10</sup>, ou par une norme. Dans ce dernier cas, on introduit un désir d'approbation sociale (ou un évitement de la désapprobation) dans la fonction d'utilité de l'agent<sup>11</sup>.

De tels modèles supposent que l'agent se conforme à la norme de manière conséquentialiste. Pour lancer la discussion, reprenons l'exemple de l'honnêteté en introduction. L'honnêteté est souvent perçue dans le choix rationnel comme une norme d'action collective (Akerlof, 1983 ; Pettit, 1990 : 735). Il n'est pas toujours immédiatement rationnel de dire la vérité, mais une société où tout le monde (ou presque) est honnête est meilleure qu'une société de menteurs. Si, de plus, on introduit un système d'approbation et de désapprobation d'autrui, on voit comment l'honnêteté peut s'avérer rationnelle après tout. Mais de quoi aurait l'air une société où la motivation à l'honnêteté serait purement conséquentialiste ? Il est difficile de croire qu'un agent se déclarant honnête dans le but explicite d'être apprécié d'autrui et d'éviter les sanctions soit digne d'approbation. En fait, l'agent maximisera son utilité en *paraissant* sincèrement honnête, donc en paraissant agir de manière non conséquentialiste. Le grain de sable dans l'engrenage d'un modèle rationnel de l'honnêteté est que l'honnêteté conséquentialiste n'est jamais un aussi bon « choix » que l'honnêteté sincère, et, néanmoins, cette première forme d'honnêteté est la seule qu'il puisse recommander.

Riker (1980) s'oppose à une conception de la confiance comme « attitude » (sans doute non conséquentialiste) pour y préférer une conception conséquentialiste :

*Dig just below the surface, even of diffuse trust or support or passive alienation, and numerous individual choices based on rational calculation will appear. The conclusion is that people do think about their circumstances, do seek their advantage, and are not precluded by personality or history or social circumstances from changing their judgments as their situations change* (Riker, 1980 : 6).

Il faut tout de même reconnaître qu'un agent dont « l'attitude » est de respecter ses promesses sera plus digne de confiance que celui qui ne les respecte que lorsque cela tourne à son avantage. La rationalité stricte de Riker ne peut rendre compte de ce phénomène.

Certains auteurs rationalistes ont reconnu le problème. Une solution proposée consiste à rendre la rationalité de l'action latente. Hardin (1995 : 86) prend l'exemple du slogan « *Black is Beautiful* » chez les Noirs américains. Même si l'adoption par chacun d'un tel slogan s'avère rationnel par l'effet de renforcement de la solidarité de groupe qui en résulte, l'adopter explicitement *pour cette raison* serait contre-productif. Il propose alors un modèle rationnel fonctionnaliste admettant la rationalité latente : le bénéfice collectif sert

10. La branche néo-institutionnaliste de la théorie du choix rationnel traite justement de ce genre de situation (Ostrom, 1986 ; Shepsle, 1989 ; Goodin, 1996).

11. Comme représentants de tels modèles, voir Höllander (1990), Frank (1998), Nee et Ingram (1998) et Bernheim (1994).

d'incitatif à l'adoption du slogan par une boucle de rétroaction, mais sans que l'agent ne s'en aperçoive. Brennan et Pettit (2000) offrent une approche similaire. L'agent adopte une stratégie rationnelle lui permettant de gagner l'estime d'autrui, mais puisque la recherche ouverte d'estime est vouée à l'échec, il découpe sa stratégie en une succession de petites actions qui feront en sorte qu'il finira par oublier sa véritable motivation. C'est ici aussi un modèle fonctionnaliste, car bien que l'agent n'agisse pas pour gagner l'estime d'autrui, si cette estime cesse alors il cessera son action. La rationalité latente constitue une hypothèse auxiliaire à l'hypothèse de rationalité, qui stipule que si la motivation rationnelle influence négativement l'utilité de l'agent, alors il est possible d'être rationnel sans être rationnellement motivé.

On retrouve, dans les modèles évolutionnaires, une autre manière de se départir des motivations rationnelles tout en conservant la rationalité. La théorie évolutionnaire du choix rationnel cherche avant tout à simplifier au maximum le concept de l'agent. Ici, les individus s'apparentent plus à des « hôtes » qu'à des « agents » ; ils acquièrent leurs préférences et leurs stratégies à travers un processus semblable à la sélection naturelle plutôt que par choix volontaire. Dans un jeu évolutionnaire, les joueurs sont tirés au hasard dans une population et placés dans le jeu dans où ils appliquent leurs stratégies et sur lequel ils n'ont à peu près pas d'information a priori ; cette procédure est ensuite répétée un certain nombre de fois, et la stratégie « gagnante » est celle qui se sera propagée le plus dans la population selon un mécanisme de sélection particulier<sup>12</sup>. Le principe général de sélection est ce que Bowles et Gintis nomment « répliation différentielle » : « *Durable aspects of behavior, including norms, habits, and rules of thumb, may be accounted for by the fact that they have been copied, diffused, and hence replicated, while other traits have not* » (Bowles, Gintis, 1998 : 212). Dans les modèles évolutionnaires, les stratégies sont *sélectionnées* par un algorithme exogène à l'agent ; elles ne sont pas *choisies* par un processus endogène ; en d'autres mots, on sort l'hypothèse de rationalité de l'agent pour la placer dans l'algorithme. Il n'y a donc même pas lieu de se demander si le choix s'avère conséquentialiste ou non. Nul besoin de préciser que ce genre d'explication s'avère encore plus fonctionnelle que les précédentes.

#### *Modèles rationnels non standard*

Certains théoriciens s'intéressent de plus près à la multiplicité des motivations des agents. Insatisfaits de la modélisation rationnelle *ad hoc* des valeurs et des normes, ils assignent des caractéristiques particulières aux préférences dites « morales » par rapport aux préférences plus matérielles, tout en voulant respecter les postulats de la théorie du choix rationnel. Dans le champ économique, nous rencontrons fréquemment des modèles de préférences multiples. Un agent peut choisir d'acheter telle voiture plutôt que telle autre sur le marché parce qu'elle représente le meilleur compromis entre ses

12. Voir Weibull (1998) pour un excellent résumé de la théorie ; également Danielson (2004).

préférences pour la performance, la consommation d'essence et l'apparence, par exemple. On qualifie ces préférences d'*homogènes*, car, partageant la même nature (ici comme caractéristiques de voitures), on peut les classer sur une même échelle globale d'utilité. L'agent peut sacrifier un peu de l'utilité que lui procure la satisfaction de la préférence A pour une plus grande satisfaction de la préférence B. La théorie du choix rationnel est très bien outillée pour aborder ce genre de décision et consacrer un ordonnancement des possibilités en assignant à chacun des choix une valeur globale d'utilité. Mais avec des préférences *hétérogènes*, où l'on cherche à distinguer conceptuellement préférences matérielles et morales, le calcul de l'utilité finale devient une affaire beaucoup plus délicate.

Nous avons retenu deux types de modélisation des préférences hétérogènes dans la littérature rationaliste. Les modèles d'*utilités multiples* établissent une distinction entre utilité égoïste et non égoïste, ce dernier terme pouvant signifier, selon le modèle, utilité altruiste, sociale, ou encore, émanant des valeurs. Les modèles de *préférences hiérarchiques*, pour leur part, conçoivent les valeurs comme des préférences de second ordre, gouvernant celles de premier ordre. Ces approches engendrent des tensions au sein de la théorie du choix rationnel, car elles remettent en question le sens même « d'utilité » et de « préférence ». Il ne faudrait pas confondre les utilités multiples avec ce que nous pourrions nommer les « utilités additives », de la forme  $U = UA + UB$ , où la fonction d'utilité nous montre ses diverses composantes et comment elles affectent la courbe d'utilité *unique* de l'agent. De tels modèles demeurent parfaitement en règle avec la théorie du choix rationnel<sup>13</sup>. Ce qui nous intéresse dans l'idée d'utilités multiples, c'est que ce sont des utilités fondamentalement conflictuelles, qui ne peuvent directement s'intégrer dans une fonction unique. Etzioni (1986 : 171-72) propose de distinguer « plaisir » et « devoir », le premier concernant les choix intéressés, et le second, le respect de nos valeurs et normes. Son raisonnement est que nous associons nos préférences égoïstes à une recherche de plaisir, alors que le devoir nous enjoint fréquemment à accomplir des tâches qui nous sont déplaisantes. Toutefois, il ne nous indique pas comment calculer un choix final fondé sur ces deux paramètres. Si on les additionne tout bonnement, nous sommes de retour à la théorie standard, et le « devoir » ne jouit alors d'aucune propriété spéciale. Si cela n'est pas désirable, alors on doit expliquer comment deux paramètres présumés incommensurables peuvent aboutir à un choix unique :

*To compete with the neoclassical paradigm of rational choice, proponents of multiple utility representations must show how conflicting motivations are*

---

13. Höllander (1990) nous en fournit un bon exemple avec son modèle de la « coopération volontaire » en groupe. Sa fonction d'utilité comprend deux éléments principaux : la contribution individuelle et l'approbation du groupe. Ici, biens matériels et approbation peuvent s'échanger librement. Son modèle d'approbation dans la production de biens publics prédit, à l'équilibre, un certain niveau de contribution et d'approbation correspondante. Cet équilibre est ensuite perçu par tous comme une norme sociale, cadrant ainsi avec le principe des attentes mutuelles de la théorie de l'échange social.

*resolved into outcomes. But as soon as this step is taken, they are involved in all of the ‘commensurability’ of extended utility representations (Isaac, 1997 : 547-548).*

Ce que Isaac nomme « utilité étendue » correspond à nos utilités additives. Examinons à présent deux exemples de modèles d'utilités multiples.

Dans le modèle de Heath (2001), les résultats des actions reçoivent une valeur d'utilité (espérée) standard, alors que les actions elles-mêmes se voient attribuer une valeur de « convenance » (*appropriateness*), une « contrainte déontique » qui vient spécifier le degré (numérique) de permission, d'interdiction, ou d'obligation de l'action. Ces contraintes déontiques proviennent des normes sociales et des valeurs individuelles. Avec  $A(a)$  pour la convenance de l'action  $a$  et  $U(a)$  pour l'utilité standard de  $a$ , nous obtenons la « fonction de valeur » globale :  $V(a) = kA(a) + U(a)$ . Ces utilités sont différenciées, car, maintient-on, l'évaluation « normative » de l'action n'est pas la même chose que l'évaluation « instrumentale » de son résultat. Mais pour que  $V(a)$  ait un sens, il faut bien que ces utilités se situent sur une même échelle (Heath, 2001 : 374), donc soient commensurables et comparables d'une manière qui demeure inexplicitée. Heath adopte une défense plutôt méthodologique qu'épistémologique :

*I think it is very important to keep distinct the contribution that norms of action and desires for outcomes make to deliberation, because some social behavior becomes intelligible only when these two are treated separately. [...] Thus any attempt to merge norms and desires into a unified utility theory function results in a loss of information that in turn renders certain aspects of social interaction needlessly obscure (Heath, 2001 : 378-79).*

Au premier regard, tout cela a l'air d'un modèle d'utilités additives, mais Heath maintient que l'évaluation normative ne peut être traitée comme un élément de l'utilité standard<sup>14</sup>. Il prétend pouvoir résoudre le Dilemme du prisonnier en attribuant une valeur de convenance élevée à la coopération, mais on peut aboutir au même résultat simplement en augmentant directement l'utilité (standard) de la coopération telle qu'on la retrouve dans la matrice de jeu. Dans la forme développée du jeu, cela reviendrait à écrire la valeur de convenance directement sur les vecteurs plutôt que sur les points terminaux, ce qui est absurde en théorie des jeux. De toute évidence, les valeurs normatives ne jouissent pas ici de caractéristiques spéciales par rapport aux utilités standard.

---

14. L'argument est que les préférences sur les manières d'agir violent la théorie Von Neumann-Morgenstern de l'utilité (face à deux loteries distinctes aux résultats escomptés identiques, l'agent est supposé indifférent), sauf si elles sont directement intégrées aux résultats des choix ; mais si on les intègre, on perd de l'information précieuse sur les motivations des agents (Heath, 2001 : 378-79). Cet argument tombe à plat lorsque l'on considère que, pour autant que les préférences sur les manières d'agir et les préférences sur les résultats escomptés demeurent formellement comparables, rien ne nous empêche de ventiler une fonction d'utilité VNM afin de montrer l'influence de ces deux facteurs d'utilité. C'est d'ailleurs de cette manière que la propension au risque est habituellement représentée.

Le modèle de la « surcharge morale » de Kuran (1998) constitue une autre tentative de séparation de valeurs d'utilités. Il recouvre les situations de choix où certaines options, sinon toutes, sont évaluées comme immorales par l'agent. Sans entrer dans les détails, la fonction d'utilité déduit la désutilité morale de l'action de son utilité « intrinsèque » ou standard. Même si Kuran est catégorique au sujet du statut particulier de la désutilité morale, soutenant que son modèle « *departs from the classical theory of individual choice, according to which one has a unitary preference ordering* » (Kuran, 1998 : 232), il apparaît nettement, de par sa fonction d'utilité, que la moralité constitue en fait un facteur d'utilité menant, au bout du compte, à un ordonnancement unique des préférences. Les valeurs correspondent ici à des « *judgments about preference orderings or about the choices that preferences have generated* » (Kuran, 1998 : 232), mais telles qu'on les retrouve dans le modèle, elles servent à assigner une utilité à chacun des choix, pas à l'ordonnancement en tant que tel. Le but du modèle est de démontrer qu'une fois un choix payant mais pas très moral effectué<sup>15</sup>, son utilité intrinsèque s'estompe peu après, alors que sa désutilité morale perdure, résultant en une « dissonance morale » et amenant des sentiments de culpabilité et de regret. Ce phénomène n'a toutefois rien à voir avec la fonction d'utilité précédant le choix. On doit faire appel à une seconde fonction, *post hoc*, où l'utilité intrinsèque sera décroissante avec le temps, et où, encore une fois, la moralité ne jouera pas un rôle conceptuellement distinct.

Les modèles des préférences hiérarchiques proposent une avenue différente. Ici, les valeurs ne sont pas intégrées dans la fonction d'utilité ; elles servent plutôt à filtrer les préférences immorales. Dans le paradigme du choix rationnel, on peut les considérer comme des préférences de préférences. Le modèle le plus connu est celui de Sen, dans son fameux article « *Rational Fools* » (1982). Parmi les motivations non égoïstes, il y distingue l'altruisme rationnel (la « sympathie ») du choix fondé sur les valeurs (« l'engagement ») de la manière suivante : « *If the knowledge of torture of others makes you sick, it is a case of sympathy ; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment* » (Sen, 1982 : 92, italique ajouté). La satisfaction ou la frustration d'une valeur n'affecte pas l'utilité de l'agent ; donc, si l'on considère que l'action morale comporte des coûts réels (temps, argent, etc.) et des bénéfices moraux n'entrant pas dans la fonction d'utilité, l'engagement constitue un choix non maximiseur (Sen, 1982 : 93), ce qui nous conduit à ce passage fréquemment cité, que l'engagement « *drives a wedge between personal choice and personal welfare* » (Sen, 1982 : 94). L'engagement fonctionne par l'ordonnancement des ordonnancements de préférences. L'agent ordonne

---

15. Kuran pense à un agent qui quitte un restaurant sans laisser de pourboire.

d'abord ses préférences de premier ordre selon certaines échelles d'utilité<sup>16</sup>, et ensuite procède à l'ordonnement de ces ordonnements en vue de la maximisation de valeurs, si l'on suppose qu'il aspire effectivement à agir moralement (Sen, 1982 : 100-101). Une telle modélisation des valeurs comme contraintes fortes ne peut fonctionner que si le méta-ordonnement est *lexicographique*, c'est-à-dire, si l'on n'a recours aux ordonnements de premier ordre que pour trancher les situations d'indifférence laissées par le méta-ordonnement. La procédure prend fin à l'instant où le modèle retient un seul choix maximal. Une procédure de décision si rigide ne semble pas pouvoir s'accommoder avec la plupart de nos choix quotidiens, là où la distinction entre moralité et intérêt personnel n'apparaît pas aussi claire. La proposition de Sen pour combler ce besoin de compromis moral est de laisser l'agent, selon la situation, court-circuiter certaines préférences au sommet de l'ordonnement final afin de lui permettre de choisir plus bas (Sen, 1982 : 100), mais cela jette un doute sur la véritable nature d'une contrainte morale que l'agent peut violer apparemment sans restrictions. Comment peut-il alors choisir de se conformer à une valeur tout en gardant à l'esprit son intérêt particulier ? Si l'on veut demeurer fidèle à la théorie du choix rationnel, l'attribution d'utilités aux valeurs, dans une fonction globale d'utilité, semble inévitable, mais on doit alors abandonner le modèle des préférences hiérarchiques.

Ansperger (1998) va tenter une réconciliation des deux méthodes. Afin de régler ce qu'il appelle, suivant Rawls, les « tensions de l'engagement », il propose que figure au sommet du méta-ordonnement au moins un ordonnancement lexicographique (c'est pour lui la condition d'existence d'un ordonnancement moral), suivi d'ordonnements moraux faibles pouvant faire l'objet de compromis, et, enfin, d'ordonnements selon l'intérêt personnel. Parmi l'ensemble résiduel de choix laissés par les ordonnancements lexicographiques, l'agent peut recourir à une fonction d'utilité où il assignera des valeurs d'utilité à ses valeurs, avec un poids correspondant à leur position dans l'ordonnement global (Ansperger, 1998 : 205-207). Si ce modèle s'avère correct, c'est-à-dire que le modèle des préférences hiérarchiques de Sen ne peut par lui-même gérer le compromis moral, et donc qu'une fonction quelconque d'utilités additives est requise (Ansperger, 1998 : 206), nous devons conclure que ce modèle n'a rien de spécial à offrir<sup>17</sup>. En effet, on pourrait le remplacer par un modèle d'utilités additives dans lequel les impératifs moraux les plus chers à l'agent seraient conçus comme des biens infiniment inélastiques (Brennan, 1989 : 200).

---

16. Par exemple, dans le modèle de la « rationalité de rôle », l'agent peut avoir un ordonnancement particulier pour chacun des rôles sociaux qu'il porte (Goodin, 1986 : 88-89). L'appréciation d'une situation en tant que mère n'est pas nécessairement la même qu'en tant qu'employée ou croyante.

17. Voir la critique similaire de Wolfelsperger (2001 : 76-77).

### Conclusion

Nous avons survolé quelques échantillons de modèles rationnels cherchant à rendre compte des motivations non conséquentialistes. Le problème central rencontré dans les modèles standard est que les motivations non conséquentialistes en elles-mêmes peuvent influencer l'issue du jeu, et donc, que leur réduction à des motivations conséquentialistes perdra de vue cet aspect. Les solutions à tendances fonctionnalistes ne nous apparaissent pas attrayantes. À notre avis un tel affaiblissement de l'hypothèse de rationalité ne sert pas la cause de la théorie du choix rationnel. Pour ce qui est des modèles faisant une place spécifique aux motivations non conséquentialistes, nous constatons qu'ils se retrouvent dans l'une ou l'autre de deux catégories. Soit que les utilités sont commensurables, ce qui signifie que l'on peut toujours réduire le modèle à un modèle d'utilités additives, pleinement compatible avec la théorie standard de l'utilité mais n'offrant aucune conception particulière des utilités morales, soit qu'elles s'avèrent incommensurables, et il apparaît alors impossible de pouvoir construire une fonction d'utilité appropriée sans violer les axiomes de la théorie. Les utilités additives conservent une fonction heuristique, elles nous permettent d'observer l'influence de divers facteurs dans la détermination de l'utilité globale, mais elles ne peuvent conceptuellement distinguer préférences matérielles et valeurs. La commensurabilité implique que toutes les utilités puissent se mesurer à l'échelle du « grand X », « *an abstract entity, without any content, neither pleasure nor consumption. It is merely the great common denominator, an X, into which all other values can be converted or by which all rank-ordering can be systematized* » (Etzioni, 1988 : 164). L'échec des modèles d'utilités multiples et des préférences hiérarchiques semble démontrer que l'on ne peut répondre à la critique en amont de l'hypothèse de rationalité en introduisant de nouveaux types de préférences sans rompre la théorie du choix rationnel.

### Deux théories alternatives de la rationalité

Après avoir examiné s'il était possible d'intégrer les comportements non conséquentialistes au sein de la rationalité instrumentale tout en préservant le sens, nous allons nous pencher sur deux théories alternatives de la rationalité qui évitent de ramener ces comportements à un moyen pour leur propre fin et leur donnent une place bien à eux. Les deux auteurs que nous proposons adoptent chacun une stratégie de prime abord opposée : Raymond Boudon cherche à *élargir* la notion de rationalité afin de pouvoir y inclure les comportements non conséquentialistes sans avoir à les réduire à un instrumentalisme quasi-tautologique, tandis que Jon Elster cherche à *restreindre* la portée de la rationalité aux comportements « réellement » conséquentialistes pour ensuite proposer une théorie hors rationalité des comportements non conséquentialistes.

## Le « modèle rationnel général » de Boudon

L'approche de Boudon consiste à ramener les comportements non conséquentialistes dans le giron de la rationalité, en élargissant la notion de « rationalité » au-delà de sa signification traditionnelle en théorie du choix rationnel. En gros, l'explication rationnelle ne consiste plus seulement à montrer que l'agent a choisi les meilleurs moyens en vue d'une fin, mais, plus largement, à montrer qu'il avait des raisons, instrumentales ou non, d'agir comme il l'a fait. Il décompose la théorie classique du choix rationnel en six postulats, en ordre d'implication ascendante : P1 (*individualisme*), P2 (*compréhension*, au sens d'intelligibilité de l'action), P3 (*rationalité*, au sens d'action fondée sur des raisons), P4 (*conséquentialisme/instrumentalisme*), P5 (*égoïsme*) et P6 (*maximisation*) (Boudon, 2003a : 19-21). Pour Boudon, la théorie du choix rationnel propose de réduire « toutes les formes de la rationalité à la rationalité instrumentale » (Boudon, 2003a : 133) ; elle « fait fausse route en prétendant accorder un statut général aux postulats du conséquentialisme et de l'égoïsme, qui ne sont pertinents que dans des cas particuliers » (Boudon, 2003a : 121). Il propose comme solution de remplacement le « Modèle rationnel général », moins restrictif, composé uniquement des postulats P1 à P3 (Boudon, 2003a : 49).

Le Modèle rationnel général comprend trois types de rationalité : instrumentale, cognitive et axiologique. La rationalité instrumentale est bien entendu représentée en réintroduisant le postulat P4. La rationalité cognitive concerne la formation des croyances. L'agent agit au nom d'une croyance X, car il a de fortes raisons<sup>18</sup> de croire que X est vrai (ou probable) dans un certain « contexte cognitif » (Boudon, 2003a : 61). Ici, dans la formation de ses croyances, l'agent ne cherche pas à maximiser son utilité, mais bien à déterminer si ses connaissances sont vraies (Boudon, 1998 : 188). Deux remarques s'imposent. D'abord, il faudrait savoir pourquoi Boudon oppose croyances instrumentales et cognitives. Dans les cas des croyances instrumentales, s'il s'agit d'une formation de croyances sans égards à la vérité, dans le seul but de satisfaire ses désirs, alors c'est une irrationalité (une forme de *wishful thinking* ou de *self-deception*). S'il s'agit plutôt d'une collecte d'information en vue d'une action instrumentale, la recherche de la vérité s'impose de façon implicite chez l'agent rationnel. Nous voyons que, dans ces deux cas, il n'y a pas de conception de la rationalité cognitive qui se distingue du choix rationnel. La seconde remarque concerne la relation entre rationalité et vérité. Boudon affirme que la rationalité cognitive permet de qualifier des croyances fausses de « rationnelles » pour autant que l'agent les croient vraies sur la base de raisons fortes. Contre la rationalité instrumentale, Boudon mentionne une « idée reçue », les associations entre

---

18. Boudon distingue les « bonnes raisons » propres à l'agent des « raisons fortes », qui, elles, sont entretenues par tous ceux qui partagent les mêmes « paramètres contextuels » (Boudon, 2003a : 160). On ne remarque pas de différences conceptuelles à proprement parler ; les raisons fortes sont simplement plus convaincantes pour l'agent que les bonnes raisons, et fournissent donc une base plus solide pour l'explication.



vérité et rationalité, et entre fausseté et irrationalité, « implicitement présentes dans bien des théories » (Boudon, 2003a : 81). Boudon ne mentionne pas d'exemples de ces théories, mais il est clair qu'il ne peut s'agir du choix rationnel qui, à ma connaissance, n'a jamais prétendu, sous quelque forme que ce soit, que les croyances devaient être objectivement vraies pour être rationnelles. Bien au contraire, les croyances au sein du choix rationnel sont *subjectives*, c'est-à-dire que l'agent doit *croire* qu'elles sont vraies autant que possible.

La rationalité axiologique opère de manière semblable à la rationalité cognitive : l'agent est axiologiquement rationnel s'il accomplit X parce qu'il a de fortes raisons de croire que X est juste ou bon. Alors que nous ne constatons pas de différences réelles entre la motivation rationnelle et la recherche de la vérité, la rationalité axiologique nous semble plus prometteuse comme alternative. Les raisons axiologiques se fondent sur le « programme de la morale », qui a pour but de promouvoir la « dignité de l'individu » ; ce programme se veut « indéfiniment déclinable » (Boudon, 2003a : 128), ce qui évite à Boudon d'avoir à imposer une moralité particulière comme rationnelle. Les raisons cognitives et axiologiques sont de nature non conséquentialiste.

Les comportements non conséquentialistes sont donc rationnels, au sens que Boudon donne à ce terme. Ce type de rationalité s'insère entre la rationalité instrumentale et l'irrationalité, telles que conçues dans la théorie classique du choix rationnel. La frontière entre la rationalité instrumentale et la rationalité cognitive/axiologique est assez nette, elle correspond à la frontière du conséquentialisme, représentée par l'ajout du postulat P4. Qu'en est-il alors de la frontière entre ce type de rationalité et l'irrationalité ? Quelle place occupent chez Boudon les comportements irrationnels ? Il semble bien, a priori, que ce soit le domaine des actions qui ne sont pas soutenues par des raisons, ou, en d'autres termes, qui ne respectent pas le postulat P3. Boudon a très peu à dire sur l'irrationalité ; il semble qu'il veuille pousser son concept de rationalité le plus loin possible, englobant un maximum de types de comportements. À la fin d'un article sur le « modèle cognitif » (le précurseur du Modèle rationnel général), il écrit : « *Irrationality should be given its rightful place. 'Traditional' and 'affective' actions also exists. Moreover, all actions rest on a ground of instincts* » (Boudon, 1998 : 200) ; et nous retrouvons ceci comme seule description du phénomène, au milieu d'une définition formelle de la rationalité : « Ainsi, d'une mère qui par énervement gifle son enfant, l'on dira : « Elle n'avait *pas de raisons* de gifler l'enfant, mais... ». Ce comportement est *compréhensible*, il n'est pas *rationnel* » (Boudon, 2003b : 194, en italique dans l'original). Afin de mieux comprendre, observons comment Boudon traite d'un phénomène particulier, les croyances « à moitié », où l'on croit à des mythes (par exemple) sans vraiment y croire, « on ne mettrait pas sa main au feu pour les défendre ; mais, dans bien des cas, on n'a guère d'intérêt pour leur réfutation. Plus : on répugne à les mettre en doute » (Boudon, 2003a : 139). C'est ce que Elster (2004) nomme « *wishful thinking* », et, pour lui, il s'agit clairement d'une irrationalité, car la formation de la croyance se voit biaisée par

le plaisir immédiat qu'elle apporte à l'agent. Boudon, quant à lui, désire éviter la « vision binaire » du couple rationnel/irrationnel en proposant de s'intéresser à une « typologie des modes de conviction ». Pour ce faire, il applique la notion de « satisfaction », au sens de Simon, aux croyances : l'agent a beau chercher la vérité, il se contente volontiers de croyances suffisamment proches. Cet affaiblissement du critère de « bonne raison » cognitive permet d'inclure dans une typologie sommaire des modes de conviction, des phénomènes tels l'idéologie, la « fausse conscience », ou la « mauvaise foi ». Ce sont de bonnes raisons, mais pas des raisons fortes (Boudon, 2003a : 139-141).

Le but du Modèle rationnel général est d'expliquer les comportements des agents en retraçant leurs raisons d'agir comme ils le font. La théorie du choix rationnel recherche une explication plus resserrée, elle cherche à rendre compte des intérêts des agents et ainsi à expliquer les comportements par le choix de moyens appropriés à la maximisation de ces intérêts. Il s'agit moins d'un postulat sur la nature humaine que d'un prérequis essentiel pour fins de modélisation. Comme nous l'avons constaté précédemment, les modélisations rationnelles impliquant les utilités de l'agent s'avèrent problématiques lorsque l'on y intègre des comportements non conséquentialistes. Boudon note avec justesse que les modèles de choix rationnel ont bien du mal à traiter du paradoxe du vote (Boudon, 2003a : 38-42). Considérer que les raisons de voter (ou pas) peuvent être de nature axiologique permet certes un plus grand réalisme, mais on ne voit pas immédiatement comment on pourrait en tirer un modèle explicatif rigoureux. La théorie de Boudon nous indique que ces autres rationalités existent, et qu'elles ont une influence sur les motivations des agents, mais, pour l'instant, cette théorie ne semble pas outillée pour la construction de modèles.

#### *Les « motivations mixtes » de Elster*

Dans la théorie de Boudon, l'irrationalité occupe une place minimale. Elle ne s'applique qu'aux comportements dénués de raisons, ce qui ne semble désigner que les actes instinctifs. Nous retrouvons chez Elster une définition beaucoup plus étoffée de l'irrationalité. Sa stratégie consiste en deux étapes : d'abord circonscrire le champ d'application de la théorie du choix rationnel en situant un certain nombre de comportements (dont les comportements non conséquentialistes) hors de la théorie, ensuite postuler une motivation supplémentaire à la motivation rationnelle, mettant en jeu les émotions. Commençons par sa conception de la rationalité. Elster subdivise le champ de la rationalité en théorie « étroite » et « large » (Elster, 1983 : chap. 1). La théorie étroite correspond *grosso modo* à la conception orthodoxe du choix rationnel, telle que nous l'avons abordée jusqu'ici, qui exige la cohérence des préférences (complètes, continues et transitives) et des croyances (non contradictoires).

La théorie étroite n'est pas suffisante pour Elster, car elle ne peut que soutenir que l'agent s'est effectivement comporté de manière rationnelle (ou non), sans se soucier qu'il soit tombé sur le bon choix après mûre réflexion,

par hasard, ou par conformisme aveugle. La théorie large de la rationalité inclut la théorie étroite et s'intéresse en plus aux mécanismes de formation des désirs et des croyances. Dans la théorie elstérienne du choix rationnel, les désirs et les croyances de l'agent doivent présenter une « histoire causale » à laquelle l'agent peut s'identifier<sup>19</sup> (Elster, 1986 : 15). La théorie large se situe quelque part entre la rationalité étroite et une théorie « du bien et du vrai » (Elster, 1983 : 15) ; elle cherche à expliquer les « bons choix » d'une manière plus enrichissante que la simple cohérence. Les croyances bien formées sont une affaire de *jugement*. Si l'acquisition d'information est coûteuse, il apparaît impossible de pouvoir former des croyances optimales par rapport à la situation, car on ne peut savoir à l'avance quelle sera la valeur d'une information que l'on ne possède pas encore. Tout ce que l'agent peut faire, c'est user de son jugement afin de déterminer s'il a acquis suffisamment d'information pertinente au choix qu'il a à effectuer<sup>20</sup>. Les désirs, quant à eux, doivent avoir été formés de façon *autonome*, c'est-à-dire libres, autant que possible, d'influences causales échappant au contrôle de l'agent, comme les préférences adaptatives au contexte, le conformisme (ou l'anti-conformisme), l'inertie, etc. En fait, sont considérés comme non autonomes les désirs auxquels l'agent ne « s'identifie » pas pleinement, pour reprendre la citation utilisée plus haut.

Les critères de jugement et d'autonomie n'ont pas de définition propre ; ils sont plutôt compris comme des qualificatifs que l'on applique respectivement aux croyances et aux désirs, à la condition qu'ils ne présentent pas certaines formes typiques d'irrationalité (Elster, 1983 : 24). Elster subdivise les comportements irrationnels en deux catégories : les irrationalités non motivées (*cold*), comme les erreurs inférentielles, et les irrationalités motivées (*hot*) (Elster, 1983 : 25-26 ; 1989 : 23-24). Ces dernières supposent que l'agent désire, la plupart du temps, agir de façon rationnelle ; donc qu'il entretient la rationalité comme motivation. À cela s'ajoute une seconde motivation, non rationnelle, les émotions. Elster soutient que les croyances et les désirs influencés par les émotions ne peuvent pas conduire à une action pleinement rationnelle. Les émotions peuvent influencer les désirs, à travers la colère, la jalousie, la honte, etc. « Agir sous le coup de l'émotion » signifie que l'agent n'est pas principalement motivé par l'atteinte réfléchie de ses objectifs, mais qu'il éprouve plutôt des désirs plus ou moins insensibles aux calculs coûts/bénéfices. Les croyances peuvent également être influencées par les émotions. Nous avons alors affaire au *wishful thinking*, lorsque la croyance est formée selon le « principe du plaisir » au détriment du « principe de réalité », ou à la *self-deception*, lorsqu'une croyance déplaisante se voit remplacée par une autre plus plaisante (Elster, 2004).

19. Cela constitue pour Elster une « idée préanalytique » (Elster, 1989 : 6).

20. Cet argument constitue pour Elster le talon d'Achille de la théorie étroite, qui réfute en quelque sorte sa propre hypothèse d'optimalité des décisions. Le « *satisficing* » de Simon constitue une tentative remarquable de réponse à ce problème.

La présence de motivations non rationnelles impose une limite à la portée explicative de la théorie du choix rationnel. Sa primauté normative n'est pas affectée, car, dans cette optique, elle ne fait que suggérer le meilleur moyen pour réaliser certaines fins. Dans la première mouture de ses travaux sur le choix rationnel, Elster suggérait que l'on complète la théorie par une théorie psychologique du comportement et une théorie sociologique des normes sociales (Elster, 1986 : 22-27 ; 1989 : 30-35). Plus tard (Elster, 1999), il ne retiendra que les motivations émotive et rationnelle, la motivation à se conformer aux normes étant désormais conçue de façon surtout émotive, sans totalement exclure la rationalité. La sociologie sert toujours à expliquer l'écologie des normes sociales, mais celles-ci n'ont plus le statut de motivations propres.

Elster n'offre pas de définition des émotions, se contentant de quelques esquisses de catégorisation<sup>21</sup>. L'aspect le plus important, selon lui, de la motivation émotionnelle par rapport à la motivation rationnelle, c'est son côté « viscéral », au sens où l'influence émotionnelle pousse l'agent à dévaloriser les conséquences de ses gestes<sup>22</sup> (Elster, 1999 : 287). On ne peut simplement considérer les émotions comme un coût psychique (ou un bénéfice) venant modifier les valeurs d'utilité des autres solutions, et, ainsi, l'ordonnement des préférences. Pour Elster, le caractère viscéral des émotions fait en sorte qu'en plus d'agir comme facteur d'évaluation des solutions, elles peuvent entraver, voire empêcher la réflexion rationnelle (Elster, 1999 : 304, 413). Sauf, peut-être, dans le cas d'une émotion très intense, l'agent est généralement motivé à la fois par sa raison et ses émotions ; c'est ce que Elster nomme les « motivations mixtes ». Il ne saurait être question « d'émotions rationnelles » dans le cadre de la théorie large de la rationalité. Si un comportement émotif représente effectivement le meilleur moyen pour en arriver à une fin donnée, la théorie étroite peut s'en accommoder, mais un tel comportement viole, du moins en partie, le critère de l'autonomie des désirs, et parfois même le recours au jugement dans la formation des croyances. L'agent veut pouvoir justifier ses gestes par des *raisons*, non seulement des impulsions<sup>23</sup>.

La théorie des motivations mixtes fait plus que simplement affirmer l'existence de comportements authentiquement non conséquentialistes, elle leur donne une caractéristique particulière. Chez Elster, le comportement motivé par des émotions ne peut pas se ramener à une préférence pour la satisfaction de telle ou telle émotion, comme cela se passe dans la théorie classique

---

21. Il soutient qu'il existe trop de confusion théorique dans ce domaine, et que de toute façon c'est l'aspect phénoménal des émotions qui l'intéresse le plus (Elster, 1999 : 241-43).

22. « *Visceral arousal is an important criterion for deciding that a state is an emotion rather than a simple belief-desire complex* » (Elster, 1999 : 247).

23. À mon avis, cela ne signifie pas que l'agent désire toujours agir de façon rationnelle. Celui-ci peut vouloir se laisser porter par des émotions positives. Mais lorsqu'il désire agir rationnellement, un comportement trop influencé par les émotions sera considéré comme un échec. L'expression « garder la tête froide », qui est en fait un conseil de réussite, illustre bien ce propos.

du choix rationnel, car les émotions entravent les capacités réflexives de l'agent. Pour pouvoir rendre compte de ce phénomène, la théorie du choix rationnel devrait concevoir une fonction d'utilité qui non seulement quantifie la satisfaction émotionnelle de l'agent, mais qui considère également que celui-ci se retrouve *moins* en mesure de se servir de cette fonction. Le problème, relevé chez Boudon, de la modélisation des comportements non conséquentialistes demeure. Elster propose de se servir de « mécanismes sociaux », « *frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences* » (Elster, 1998 : 45). Il apparaît évident, dans les deux cas, que prendre les comportements non conséquentialistes au sérieux implique un abandon de la rigueur de la théorie du choix rationnel.

### Conclusion

Boudon et Elster partagent une préoccupation semblable pour les comportements non conséquentialistes. Là où ils ne semblent pas s'accorder, c'est sur la conception de la rationalité. Alors que pour Boudon il suffit que l'action soit fondée en raison pour être qualifiée de rationnelle, Elster s'en tient plutôt à la rationalité instrumentale, augmentée de critères concernant la formation des désirs et des croyances. Donc, quant aux comportements non conséquentialistes qui ne sont pas *manifestement* irrationnels sous n'importe quelle définition, Boudon aura tendance à les classer comme rationnels, et Elster comme irrationnels. Bien que Boudon reconnaisse du bout des lèvres la nécessité d'une conception de l'irrationalité, il en minimise simultanément la pertinence : « Souligner la dimension cognitive de la rationalité [...] [c]'est aussi se donner les moyens d'échapper aux délimitations arbitraires entre rationalité et irrationalité, comme celle qui oppose l'irrationalité de l'*explication par les normes* à une rationalité réduite à la rationalité instrumentale [...] » (Boudon, 2003a : 160 ; en italique dans l'original). L'exemple qu'offre Boudon est frappant, car il s'agit, sans la nommer, de la position d'Elster !

L'avantage principal, selon nous, de la théorie d'Elster est qu'elle nous fournit une caractérisation d'un type de comportement non conséquentialiste, celui impliquant les émotions, qui se situe conceptuellement en porte-à-faux avec l'hypothèse de rationalité, puisque cette caractérisation remet en question la capacité de l'agent à agir de façon pleinement rationnelle. En fondant la propriété motivationnelle des valeurs sur les émotions, on peut éviter de parler « d'impératifs » suivi d'exceptions, comme le fait Etzioni, et parler plutôt d'un amalgame émotion-rationalité où les proportions peuvent varier selon la situation. La caractérisation des comportements non conséquentialistes chez Boudon nous apparaît confuse. Comme argument soutenant sa théorie en opposition à la théorie du choix rationnel, Boudon propose son propre modèle du comportement des étudiants dans le système scolaire, qu'il introduit ainsi : « *Far from making the individual educational decisions a mere effect of cost-benefit calculations, I introduced the idea that they derive, rather, from a system of*

*contextualized arguments* » (Boudon, 1998 : 193). Il nous présente ensuite le processus décisionnel de trois étudiants idéaux typiques, qui varie selon leur croyance en leurs habiletés scolaires (croyances cognitives) et leurs préférences pour un statut social élevé (croyances axiologiques). Il apparaît clairement, toutefois, que leur décision de poursuivre ou non leurs études est de nature conséquentialiste : à partir d'une préférence pour un niveau donné de statut social et d'une croyance sur ses habiletés scolaires, l'agent choisit le moyen (poursuivre ou non ses études) qui maximisera la préférence. Pourquoi alors Boudon prétend-il qu'il s'agit d'une critique du choix rationnel ? Au moins deux réponses sont possibles. D'abord, il s'en prend peut-être au postulat de l'égoïsme (P5), car pour lui le statut social n'est pas nécessairement une affaire d'intérêt personnel ; l'agent peut se révéler sensible à la réputation familiale, par exemple. Cette critique est invalide, car, dans la théorie du choix rationnel, ce à quoi l'on réfère souvent comme « égoïsme » ne signifie ni plus ni moins que l'utilité propre à l'agent ; cela n'exclut pas qu'il puisse tirer son utilité de la satisfaction d'autrui. Une seconde réponse serait qu'il ne situe pas sa critique au niveau de la décision finale, mais exclusivement au niveau des états mentaux : ces croyances ne sont pas « instrumentales », car elles concernent, respectivement, une réalité empirique et un jugement de valeur. Cette conception des rationalités cognitive et axiologique revient souvent chez Boudon ; elle demeure extrêmement problématique notamment, car, comme nous l'avons relevé plus haut, la notion de « croyance instrumentale » n'a pas de sens en dehors du *wishful thinking* et ne fait certainement pas partie de l'hypothèse de rationalité instrumentale.

Bien que Elster nous offre une distinction beaucoup plus nette entre comportements conséquentialistes et non conséquentialistes, une grande ambiguïté demeure. Ses concepts de désirs « autonomes » et de croyances issues d'un « jugement » au sein de la théorie large de la rationalité demeurent vagues et intuitifs, et sa définition des émotions ne va pas beaucoup plus loin qu'un exposé phénoménal. Néanmoins, son opposition conceptuelle entre émotions et rationalité ouvre la voie à une possibilité de modélisation combinant la rigueur de la théorie du choix rationnel et certains mécanismes sociaux traitant des motivations non conséquentialistes.

#### 1.4 Conclusion

En étudiant les modélisations rationnelles du conformisme aux valeurs et aux normes, nous en sommes venu au constat qu'au sein de la théorie du choix rationnel on n'a d'autre choix que de réduire les motivations non conséquentialistes aux motivations conséquentialistes. Si l'on adopte la stratégie de la rationalité postulée, on se prive de comportements réels pouvant avoir un effet significatif sur le phénomène collectif à l'étude, alors que si l'on adopte la stratégie dite « non standard », on a tendance à affaiblir la théorie en l'affligeant d'hypothèses auxiliaires sur l'agent. D'où l'intérêt principal de restreindre la portée de la théorie du choix rationnel aux champs et sous-champs où les

agents se montrent conséquentialistes : on préserve ainsi pleinement l'hypothèse de rationalité, la pierre de touche de la théorie, tout en produisant des modèles plausibles d'action collective.

Loin de nous l'idée de vouloir diviser les sciences sociales en champs « rationnels » et « non rationnels » et de confiner la théorie du choix rationnel à ces premiers. Nous préconisons plutôt des approches où l'explication d'un phénomène collectif peut comporter à la fois des aspects rationnels et non rationnels. C'est précisément ce qu'ont tenté d'accomplir Boudon et Elster, dans des optiques différentes. De telles approches exigent, par contre, que l'on doive abandonner la rigueur et la simplicité des modèles rationnels. En définitive, nous avons retenu la théorie des motivations mixtes d'Elster, non pas pour sa rigueur, car elle nous semble encore à un stade précoce, mais parce que ses propositions sur l'irrationalité et les émotions nous semblent une voie prometteuse pour une théorie qui permettra de traiter à la fois des comportements conséquentialistes et non conséquentialistes.

Nous pourrions faire le même constat à propos d'autres théories sociales, comme, par exemple, la théorie de la démocratie délibérative (Gagné, 2002). Ici, les citoyens sont supposés ne pas chercher à maximiser leur intérêt personnel, et ils sont également supposés être ouverts aux arguments d'autrui en toute bonne foi ; c'est ce que nous pourrions nommer « l'hypothèse de l'agent raisonnable », parallèlement à l'hypothèse de rationalité. Mais que faire lorsque les agents réels ne s'avèrent pas raisonnables ? Encore une fois, au lieu de chercher à « sauver » la théorie en excluant certains comportements et en en modifiant d'autres, il vaudrait mieux, selon nous, adopter une approche mixte de motivations, en invoquant les modèles délibératifs là où ils se révèlent vraiment pertinents et en n'hésitant pas à faire appel à la rationalité dans d'autres cas.

## Bibliographie

- Akerlof, George A. « Loyalty Filters », *The American Economic Review*, vol. 73, 1983, p. 54-63.
- Arnsperger, Christian. « Engagement moral et optimisation individuelle », in Mahieu, F.-R. ; Rapoport, H. (dir), *Altruisme. Analyses économiques*, Paris, Economica, 1998, p. 191-214.
- Becker, Gary. « The Economic Approach to Human Behavior », in Elster, J. (dir), *Rational Choice*, New York Univ. Press, 1986, p. 108-122.
- Bernheim, B. Douglas. « A Theory of Conformity », *Journal of Political Economy*, vol. 102, 1994, p. 841-877.
- Boudon, Raymond. « Social Mechanisms Without Black Boxes », in Hedström, P. ; Swedberg, R. (dir), *Social Mechanisms*, New York, Cambridge Univ. Press, 1998, p. 172-203.
- . *Raison, bonnes raisons*, Paris, PUF, 2003a.
- . « Rationalité », in Boudon, R. et al. (dir.), *Dictionnaire de sociologie*, Paris, Larousse, 2003b, p. 194-195.

- Bowles, Samuel ; Gintis, Herbert. « How Communities Govern : The Structural Basis of Prosocial Norms », in Ben-Ner, A. ; Putterman, L. (dir), *Economics, Values, and Organization*, New York, Cambridge Univ. Press, 1998, p. 206-230.
- Brennan, Geoffrey ; Pettit, Philip. « The Hidden Economy of Esteem », *Economics and Philosophy*, vol. 16, 2000, p. 77-98.
- Brennan, Timothy J. « A Methodological Assessment of Multiple Utility Frameworks », *Economics and Philosophy*, vol. 5, 1989, p. 189-208.
- . « The Trouble with Norms », in Koford, K. J. ; Miller, J. B. (dir), *Social Norms and Economic Institutions*, Ann Arbor, Univ. of Michigan Press, 1991, p. 85-94.
- Danielson, Peter. « Rationality and Evolution », in Mele, A. R. (dir), *The Oxford Handbook of Rationality*, Oxford Univ. Press, 2004, p. 417-437.
- Demeulenaere, Pierre. *Homo oeconomicus*, Paris, PUF, 1996.
- Elster, Jon. *Sour Grapes*, New York, Cambridge Univ. Press, 1983.
- . « Introduction », in Elster, J. (dir), *Rational Choice*, New York Univ. Press, 1986, p. 1-33.
- . *Solomonic Judgments*, New York, Cambridge Univ. Press, 1989.
- . « A Plea for Mechanisms », in Hedström, P. ; Swedberg, R. (dir), *Social Mechanisms*, New York, Cambridge Univ. Press, 1998, p. 45-73.
- . *Alchemies of the Mind*, New York, Cambridge Univ. Press, 1999.
- . « Costs and Constraints in the Economy of the Mind », in Brocas, I. ; Carrillo, J. D. (dir), *The Psychology of Economic Decisions, vol. 2, Reasons and Choices*, New York, Oxford Univ. Press, 2004, p. 3-14.
- Etzioni, Amitai. « The Case for a Multiple-Utility Conception », *Economics and Philosophy*, vol. 2, 1986, p. 159-183.
- Etzioni, Amitai. *The Moral Dimension*, New York, Free Press, 1988.
- Ferejohn, John. « Rational Choice Theory and Social Explanation », *Economics and Philosophy*, vol. 18, 2002, p. 211-234.
- Follesdal, Dagfinn. « The Status of Rationality Assumptions in Interpretation and in the Explanation of Action », *Dialectica*, vol. 36, 1982, p. 301-316.
- Frank, Robert H. « Social Norms as Positional Arms Control Agreements », in Ben-Ner, A. ; Putterman, L. (dir), *Economics, Values, and Organization*, New York, Cambridge Univ. Press, 1998, p. 275-295.
- Gagné, Leary. « La délibération circonstancielle en théorie démocratique », *Philosophiques*, vol. 29, 2002, p. 327-350.
- Goldthorpe, John H. *On Sociology*, New York, Oxford Univ. Press, 2000.
- Goodin, Robert E. « Laundering Preferences », in Elster, J. ; Hylland A. (dir), *Foundations of Social Choice Theory*, New York, Cambridge Univ. Press, 1986, p. 75-102.
- . « Institutions and Their Design », in Goodin, R.E. (dir), *The Theory of Institutional Design*, New York, Cambridge Univ. Press, 1996, p. 1-53.
- Hardin, Russel. *One for All: The Logic of Group Conflict*, Princeton Univ. Press, 1995.
- Hausman, Daniel M. ; McPherson, Michael S. « Taking Ethics Seriously », *Journal of Economic Literature*, vol. 31, 1993, p. 671-731.
- . *The Inexact and Separate Science of Economics*, Cambridge (UK), Cambridge Univ. Press, 1992.
- Heath, Joseph. « Rational Choice with Deontic Constraints », *Canadian Journal of Philosophy*, vol. 31, 2001, p. 361-388.



- Höllander, Heinz. « A Social Exchange Approach to Voluntary Cooperation », *The American Economic Review*, vol. 80, 1990, p. 1157-1167.
- Isaac, Alan G. « Morality, Maximization, and Economic Behavior », *Southern Economic Journal*, vol. 63, 1997, p. 559-570.
- Kuran, Timur. « Moral Overload and its Alleviation », in Ben-Ner, A. ; Putterman, L. (dir), *Economics, Values, and Organization*, New York, Cambridge Univ. Press, p. 231-266, 1998.
- Levitt, Steven D. ; Dubner, Stephen J. *Freakonomics*, New York, William Morrow, 2005.
- Nee, Victor ; Ingram, Paul. « Embeddedness and Beyond : Institutions, Exchange, and Social Structure », in Nee, V. ; Brinton, M. C. (dir), *The New Institutionalism in Sociology*, New York, Russel Sage Foundation, 1998, p. 19-45.
- Ostrom, Elinor. « An Agenda for the Study of Institutions », *Public Choice*, vol. 48, 1986, p. 3-25.
- Pettit, Philip. « *Virtus Normativa* : Rational Perspectives », *Ethics*, vol. 100, 1990, p. 725-755.
- . « The Virtual Reality of *Homo Economicus* », *The Monist*, vol. 78, 1995, p. 308-329.
- Riker, William H. « Political Trust as Rational Choice », in Lewin, L. ; Vedung, E. (dir), *Politics as Rational Action*, Dordrecht, D. Reidel, 1980.
- Satz, Debra ; Ferejohn, John. « Rational Choice and Social Theory », *Journal of Philosophy*, vol. 91, 1994, p. 71-87.
- Sen, Amartya. « Rational Fools », in Sen, A. (dir), *Choice, Welfare, and Measurement*, Cambridge (MA), MIT Press, 1982, p. 84-106.
- Shepsle, Kenneth A. « Studying Institutions : Some Lessons From the Rational Choice Approach », *Journal of Theoretical Politics*, vol. 1, n° 2, 1989, p. 131-47.
- Weibull, Jörgen W. « Evolution, Rationality and Equilibrium in Games », *European Economic Review*, vol. 42, 1998, p. 641-649.
- Wolfelsperger, Alain. « La modélisation économique de la rationalité axiologique. Des sentiments moraux aux mécanismes sociaux de la moralité », in Boudon, R. ; Demeulenaere, P. ; Viale, R. (dir), *L'explication des normes sociales*, Paris, PUF, 2001, p. 63-92.