

# Une théorie cognitive de la rationalité axiologique<sup>1</sup>

Nicolas Gravel<sup>2</sup>      Emmanuel Picavet<sup>3</sup>

le 19 mars 1999

<sup>1</sup>Ce travail a été présenté dans le cadre du séminaire “Le mental et le social”, à l’Institut d’histoire et de philosophie des sciences et des techniques, le 20 janvier 1999. Le second auteur remercie le Centre Economie et Philosophie du THEMA pour son accueil au cours des recherches conduites en 1998.

<sup>2</sup>THEMA, UFR d’économie et de gestion, Université de Cergy-Pontoise, 33, Bd. du port, 95 011 Cergy-Pontoise CEDEX.

<sup>3</sup>IHPST, UFR de philosophie, Université Panthéon - Sorbonne, 17 rue de la Sorbonne, 75231 Paris CEDEX 05.

### **Abstract**

Cet article propose une explication des phénomènes d'adhésion à des normes dans le cadre d'une méthodologie individualiste. L'adhésion à des normes y est envisagée comme le résultat d'un processus rationnel de cognition pour des individus impliqués dans une situation d'interaction sociale. Un modèle d'explication des normes est proposé pour une classe de situations d'interaction sociale où les acteurs ont des préférences spécifiques connues. La généralisation du modèle à des situations où les préférences des agents ne sont pas déterminés est également esquissée.

# 1 Introduction

L'objectif de cet article est d'expliquer, dans une perspective individualiste et conséquentialiste, l'acceptation fréquente de certaines *normes* de comportement qui, à première vue du moins, ne semblent pas pouvoir se justifier aisément dans cette perspective. Ces normes de comportement prescrivent des conduites individuelles qui paraissent incompatibles avec la poursuite de l'intérêt individuel de l'acteur et sont néanmoins compréhensibles en termes de choix individuels raisonnés. En voici quelques exemples.

1) L'actualité sportive révèle de temps à autre une réprobation répandue, chez les sportifs eux-mêmes, du dopage dans la compétition de haut niveau. Comment un sportif peut-il réprover une pratique individuelle qui, si sa disposition subjective à sacrifier son capital de santé pour augmenter ses chances de gains est élevée, est celle qui lui permet le mieux d'atteindre son objectif ?

2) De nombreux électeurs des démocraties occidentales votent alors que, si l'on excepte le plaisir intrinsèque qu'ils retirent peut être du fait de se rendre aux urnes, ils n'ont aucun intérêt à le faire car la probabilité que leur vote individuel influence le résultat de l'élection est négligeable.<sup>1</sup>

3) Certains promeneurs s'abstiennent de jeter leurs détritus dans les lieux publics malgré la gêne individuelle que ce comportement leur impose et ce, même lorsque le risque de sanction que ce comportement vertueux permet d'éviter est virtuellement nul.

4) De nombreuses personnes témoignent (parfois au péril de leur vie) de leur attachement à certains droits "fondamentaux" de la personne humaine d'une manière qui ne semble pas directement subordonnée à l'existence d'avantages personnels liés, pour eux, au respect de ces droits.

Dire que nous voulons *expliquer* de telles normes de comportement dans l'optique d'une méthodologie *individualiste* et *conséquentialiste* appelle quelques précisions. L'aspect 'individualiste' de l'entreprise est, peut être le plus aisé à justifier dans ce contexte. Chacun sait en effet que, dans les sciences humaines, l'individualisme méthodologique désigne le schéma général d'appréhension des faits sociaux (le taux de suicide chez les jeunes, la baisse des prix des ordinateurs, etc.) en tant que résultats de l'agrégation des comportements individuels impliqués dans la définition de ces faits. Mais l'individualisme méthodologique serait une bien plate tautologie s'il se limitait à ce banal énoncé de principe. Après tout, affirmer que le taux de suicide chez les jeunes est de 2% parce que 2% des jeunes individus choisissent de mettre fin à leur jour ne fait pas progresser significativement notre connaissance des phénomènes considérés. Si l'on n'explique pas *pourquoi* les individus concernés se sont comportés de la manière qui donne lieu au phénomène social en question, on vide l'individualisme méthodologique de tout intérêt explicatif.

Le moyen usuellement employé dans les sciences humaines pour rendre opérationnel l'individualisme méthodologique est de prêter une certaine *rationalité* aux comportements individuels. Nous aurons l'occasion de faire rapidement le point sur les différents sens que l'on peut donner à ce terme. Contentons nous à ce stade de qualifier de rationnelle toute action qui, parmi les actions disponibles à l'acteur, apparaît comme la meilleure du point de vue de la capacité à satisfaire un objectif particulier.

---

<sup>1</sup>Sur ce "paradoxe du vote" voir Ferejohn et Fiorina [17], Boudon ([8], ch. 2) et les références données dans Green et Shapiro [27].

Dans cette optique, le projet d'*expliquer* des normes individuelles de comportement telles que celles que nous avons énumérées dans les exemples ci-dessus revient à identifier et à préciser les principes de rationalité de ces normes qui *pourraient* justifier leur adoption par des individus. L'objectif que nous poursuivons est en effet d'identifier les propriétés de rationalité que possèdent certaines normes de comportement, et qui sont susceptibles d'expliquer leur sailance. Mais notre approche ne permet pas de prévoir si un individu adoptera une norme satisfaisant ces propriétés plutôt qu'une autre. La théorie que nous examinons dans cet article n'est pas une théorie psychologique. Elle ne fournit aucun moyen de résoudre les dilemmes que peuvent vivre les acteurs entre l'exigence de la morale et celle de l'intérêt personnel unilatéral. Elle ne fait que donner une définition plausible de ce qui vaut comme "la morale", destinée à s'ajouter à celle, plus complète et mieux comprise, de ce qui passe pour "l'intérêt". Par exemple, les propriétés que nous examinerons nous conduiront à juger "rationnelle" l'adoption de la norme suivant laquelle il ne faut consommer aucun produit dopant dans un entraînement sportif de haut niveau. Mais le fait qu'il y ait de bonnes raisons de réprouver le dopage n'est en rien contradictoire avec la persistance de la consommation individuelle de produits dopants, ni même avec la rationalité individuelle d'une telle consommation du point de vue de l'intérêt personnel unilatéral de l'acteur. Cette démarche s'inscrit dans la lignée d'une série de travaux récents en sociologie (par exemple, Boudon ([5], [6], [8]), Coleman [10], Demeulenaere ([13], [14], [15]), Fillieule ([18], [19], [20]) et Ostrom [41]) qui visent à approfondir la compréhension de la rationalité individuelle pour rendre compte des normes.

Deux caractéristiques distinguent, à notre avis, notre approche d'autres tentatives "individualistes" d'explication des normes. La première est la grande importance accordée à la *connaissance* que développent les acteurs individuels de l'environnement à l'intérieur duquel ils agissent. La justification des normes que nous proposons résulte de ce processus *cognitif* bien décrit par Giddens [23] (p. 67) lorsqu'il affirme que "les êtres humains produisent des théories de leurs actions". Dans ce processus, nous mettons en relief les éléments qui ont été privilégiés dans les sciences humaines, et qui ont été précisément identifiés dans la théorie des jeux moderne, à savoir: les actions disponibles pour les agents, les conséquences que l'adoption simultanée de ces actions peut avoir eu égard aux objectifs poursuivis et la reconnaissance mutuelle de la qualité d'agent rationnel (ce que les théoriciens des jeux connaissent sous l'expression de connaissance commune de la rationalité).<sup>2</sup> C'est pour cette raison que nous qualifions notre démarche de 'cognitivist', même si le sens que nous donnons à ce terme diffère de celui qui sous-tend son emploi chez de nombreux autres auteurs.<sup>3</sup>

La seconde particularité de notre approche est de tenter de justifier les normes par le caractère favorable des *conséquences* que leur adoption conjointe pourrait avoir pour les participants. En ce sens, nous ne postulons aucune opposition de principe entre les prescriptions d'une norme de comportement et l'intérêt individuel des agents auxquels cette norme est prescrite. Dans cette perspective, il est essentiel de préciser ce que l'on entend par "conséquences favorables". Il faut aussi spécifier la manière dont les conséquences dépendent des actions individuelles.

<sup>2</sup>Voir Bernheim [3] pour une exploration systématique de cette hypothèse.

<sup>3</sup>Par exemple, celui que lui donnent les travaux de psychologie cognitive comme ceux de Kahneman et Tversky [30].

Le critère de rationalité des normes que nous proposons mobilise un principe qui a été abondamment commenté en philosophie sous des formes diverses, dans la tradition kantienne en particulier : la nécessité de procéder à une opération d'*universalisation* des motifs d'actions individuelles afin d'en évaluer la rationalité.<sup>4</sup> Comme nous essaierons de l'expliquer, une telle opération s'impose, indépendamment de l'adoption d'un système éthique particulier (fût-il universaliste comme l'éthique kantienne), à titre d'exigence *minimale* de rationalité cognitive de la part d'un individu qui participe à une interaction sociale et qui comprend, dans ce contexte, que la conséquence de son choix dépend du choix des autres participants. Cette interdépendance des participants à une interaction sociale rend inappropriée la formulation unilatérale d'un critère de rationalité de l'action individuelle. Dans une interaction sociale la rationalité d'un principe d'action doit nécessairement être celle d'une *norme sociale de comportements individuels*. Comme nous essaierons de l'expliquer, une telle norme doit prescrire aux individus des comportements qui peuvent être reconnus comme *analogues* par eux (au sens où ils relèvent d'une motivation semblable) tout en étant, sous cette contrainte d'analogie, rationnels relativement à leur propres objectifs individuels. A cet égard, notre démarche se rapproche de travaux récents en économie, ayant pour but la formulation de critères particuliers pour les normes de comportement, à partir d'opérations d'universalisation.<sup>5</sup>

Nous pensons que le modèle proposé parvient à expliquer l'importance accordée à certaines normes dans des situations d'interaction sociale où les participants sont animés de motivations bien spécifiques. Les exemples de normes de non-dopage, de vote, ou de respect de l'environnement mentionnés plus haut appartiennent à cette catégorie. Par contre, le modèle ne permet pas de justifier des normes comme le respect des droits fondamentaux de la personne, à laquelle sont attachés de nombreux individus, et qui forment une partie de l'*éthos* des sociétés démocratiques. Les problèmes que posent ces normes à une approche individualiste "rationaliste" tiennent au fait qu'elles sont, en général, supposées valables indépendamment des préférences que peuvent avoir les acteurs. Par exemple, la norme suivant laquelle on ne doit pas empêcher un individu de pratiquer de manière "discrète" certains rites religieux, du moins telle qu'elle est usuellement défendue, est réputée valable quels que soient les objectifs et les préférences des individus. Nous dirons quelques mots de la manière dont les normes de respect des droits pourraient être appréhendées dans le prolongement de l'analyse des situations dans lesquelles les acteurs sont animés de motivations spécifiques. Nous ne proposerons par contre aucun critère opérationnel autorisant une sélection précise de certaines de ces normes.

Le plan de l'article est le suivant. Dans la seconde section, nous introduisons un modèle de la rationalité axiologique, ou rationalité de l'adoption des normes, en le présentant comme un cas particulier, applicable dans les situations

---

<sup>4</sup>On pense évidemment à la formule célèbre de Kant ([32], p. 103) suivant laquelle "je dois toujours me conduire de telle sorte que je puisse aussi vouloir que ma maxime devienne une loi universelle". Voir également Sidgwick [59] pour une perspective utilitariste sur l'universalisation.

<sup>5</sup>Une liste non-exhaustive de ces travaux pourraient inclure Laffont [36], Guttman [28], Sugden [60], Etzioni [16], Bordinon[4], Roemer ([50],ch.6) et Bilodeau et Gravel [?]. On trouve une discussion critique de ce type d'approche dans Wolfelsperger [62]. Une application du critère d'universalisation dans le contexte de la théorie des choix collectifs est proposée par Gravel, Laslier et Trannoy [?]. La thématique de l'universalisation recoupe partiellement celle de la réciprocité, explorée dans Kolm [34].

d'interaction sociale, d'une conception plus générale, et somme toute traditionnelle, de la rationalité. Dans la troisième partie, nous définirons précisément et formellement notre théorie de la rationalité axiologique dans le contexte des jeux non-coopératifs représentés en forme normale et nous montrerons au travers d'exemples comment elle permet d'identifier un certain nombre de normes communément reconnues comme importantes. La quatrième partie sera consacrée à une brève discussion du problème que pose la justification de normes imposant le respect des droits. La conclusion mentionnera quelques problèmes ouverts.

## 2 De la rationalité instrumentale à la rationalité axiologique

Nous avons évoqué en introduction la *rationalité* attribuée aux comportements individuels. Qu'entend-t-on précisément par ce terme? Depuis longtemps en sciences économiques (voir par exemple Walliser [61]), et de manière explicite en sociologie depuis les travaux récents de Raymond Boudon (voir par exemple [5]), on distingue au moins deux grandes acceptions de cette notion, que nous examinons tour à tour.<sup>6</sup>

### 2.1 La rationalité instrumentale

Selon la première de ces conceptions, on considère comme Rationnelle toute action qui, parmi les actions disponibles pour l'agent, apparaît comme la meilleure du point de vue de la capacité à satisfaire un *objectif* particulier doté d'une *structure* minimale. Cet "objectif" peut lui-même prendre des formes très variées, allant du profit réalisé par un propriétaire d'entreprise aux préférences subjectives d'un consommateur pour les paniers de biens qu'il peut se procurer en passant par le nombre de voix que peut espérer obtenir un candidat à une élection. Quel que soit cet objectif, la théorie la plus générale exige *au moins* qu'il permette un classement *transitif* et *complet* des actions possibles.<sup>7</sup> Il peut évidemment, comme dans le cas du profit et du nombre de voix, prendre une forme plus structurée que le simple classement complet et transitif. Mais s'il ne produit pas au moins un tel classement, l'objectif de l'agent ne peut donner lieu à un comportement de choix rationnel.<sup>8</sup>

Bien que très générale, parce que ne préjugant aucunement de la nature de l'objectif de référence (pourvu que celui-ci soit complet et transitif), cette définition de la rationalité n'est pas tautologique. Tout comportement n'est pas nécessairement rationnel par rapport à un objectif. Pour qu'il le soit, le comportement doit satisfaire des propriétés particulières potentiellement réfutables

---

<sup>6</sup>Sur l'insertion de cette problématique dans l'évolution du débat sur l'individualisme méthodologique, nous renvoyons le lecteur à Boudon, Bouvier et Chazel [9] et à Goldthorpe [24].

<sup>7</sup>Un critère de comparaison d'actions est transitif si, pour toutes actions  $a$ ,  $b$  et  $c$ , il considère  $a$  au moins aussi bonne que  $c$  dès lors qu'il considère  $a$  au moins aussi bonne que  $b$  et  $b$  au moins aussi bonne que  $c$ . Un critère est complet s'il permet toujours la comparaison de deux actions entre elles.

<sup>8</sup>En toute rigueur, c'est la propriété d'acyclicité (moins exigeante que la transitivité) qui est nécessaire pour donner lieu à un comportement de choix rationnel (tout au moins dans un environnement où l'individu a le choix entre des actions en nombre fini). Voir Sen ([53], ch. 1\*) pour une discussion formelle de ces propriétés.

comme, par exemple, les axiomes de la préférence révélée en théorie du consommateur ou, dans des contextes plus abstraits, les propriétés  $\alpha$  et  $\beta$  de Sen.<sup>9</sup> Pour cette raison, cette première conception “*instrumentale*” de la rationalité est bien dotée d’un réel pouvoir de discrimination relativement aux faits sociaux. Il est théoriquement possible de retrouver l’objectif que vise à poursuivre le comportement rationnel d’un acteur à partir de la seule observation de ce comportement. Sous une hypothèse de stabilité de l’objectif, il est alors possible d’utiliser cet objectif pour prévoir le comportement de l’acteur dans des contextes nouveaux. En somme, la conception instrumentale de la rationalité individuelle permet d’expliquer et de prévoir d’une manière significative le comportement individuel.

## 2.2 La rationalité cognitive

Il est cependant juste de reconnaître que, si elle n’est pas nulle, La capacité prédictive et explicative de la conception instrumentale de la rationalité exprimée sous cette forme abstraite reste limitée. Cette limitation apparaît particulièrement éclatante dans les situations qui permettent de distinguer l’*action* d’un individu des *conséquences* que celle-ci peut avoir. Un exemple simple d’une telle situation est la décision individuelle d’achat d’un billet de loterie. L’individu qui choisit d’acheter ce billet (plutôt que de garder en sa possession le montant d’argent correspondant au prix du billet) n’a vraisemblablement pas de préférences intrinsèques pour le morceau de papier qu’il achète. Sa décision d’achat semble plutôt motivée par la conséquence monétaire que cette décision peut avoir: un gain important si le billet est gagnant et une perte modeste (correspondant au prix du billet) si celui-ci est perdant. Si le mécanisme qui relie les actions individuelles à leurs conséquences est une fonction, c’est à dire s’il ne fait correspondre à chaque action qu’une seule conséquence, la distinction entre action et conséquence n’est pas fondamentale. Mais si, comme dans l’exemple du billet de loterie, plusieurs conséquences différentes (le gain et la perte) peuvent être associées à une même action (l’achat d’un billet de loterie), la distinction devient significative. Pour comparer les actions entre elles, l’acteur individuel doit préalablement évaluer les *possibilités de réalisation* des différentes conséquences que chaque action peut avoir. Cette évaluation résulte de la *compréhension* que peut acquérir l’individu du fonctionnement de l’environnement dans lequel il se trouve. En explicitant le processus par lequel l’individu acquiert cette compréhension, on passe de la rationalité *instrumentale* à ce que l’on peut appeler la rationalité *cognitive*.

Plus précisément, le terme de rationalité cognitive qualifie le processus par lequel l’individu acquiert une connaissance de son environnement qui lui permet d’évaluer les possibilités qu’ont les différentes conséquences de son action de se réaliser. Dans cette optique, l’objet de la connaissance est le mécanisme qui associe à chaque action les différentes conséquences que celle-ci peut avoir. Cette connaissance est mobilisée par l’agent pour déterminer la meilleure conduite à tenir (du point de vue de son objectif propre).

Dans l’exemple de la loterie, le mécanisme est aléatoire et instantané. Il correspond exactement au mécanisme physique par lequel le numéro gagnant du billet est sélectionné (tirage d’une boule dans une urne, rotation d’une roue,

---

<sup>9</sup>Voir par exemple Sen [54], Samuelson [51] et Richter [49].

etc.). Dans ce contexte d'incertitude probabilisable, on fait souvent l'hypothèse que l'acheteur du billet de loterie connaît exactement les probabilités de gains et de pertes de ce billet et donc le mécanisme aléatoire du tirage du numéro gagnant. Le modèle dominant de rationalité cognitive dans ce contexte est celui de Von Neumann-Morgenstern [40] suivant lequel le décideur compare les différentes actions (par exemple acheter ou ne pas acheter le billet de loterie) sur la base de l'utilité espérée, ou moyenne, qu'il retirera de leurs conséquences.<sup>10</sup> Dans certains problèmes de décision individuelle, comme par exemple celui que pose à une entreprise le choix d'un employé sur lequel elle dispose d'une information imparfaite, il n'est cependant pas naturel de supposer que le mécanisme qui détermine les conséquences d'une certaine action puisse être probabilisé de manière objective. Qui a jamais pu être en mesure de connaître exactement la probabilité numérique que la personne que l'on envisage d'embaucher se révèle, *ex post*, incapable de satisfaire aux exigences de l'emploi auquel elle est destinée ?

Le modèle de rationalité cognitive le plus fréquemment utilisé pour appréhender des choix individuels en situation d'incertitude non probabilisable est celui de Savage [52]. Le cadre savagien conduit, comme celui de Von Neumann et Morgenstern, à un critère de classement des actions sur la base de l'espérance mathématique de l'utilité de leurs conséquences. Mais les probabilités qui interviennent dans le calcul de cette espérance mathématique ne reçoivent pas la même interprétation dans les deux modèles. Chez Von Neumann-Morgenstern, ces probabilités sont objectives et sont explicitement perçues par l'agent (par exemple, il y a une chance sur six qu'un dé à six faces de poids bien réparti, jeté sur une table, tombe sur 1). Chez Savage, les probabilités sont subjectives et reflètent des jugements de l'individu sur les possibilités de réalisation des différentes conséquences de ses actions - jugements dont l'individu peut d'ailleurs ne pas avoir conscience. L'individu rationnel du modèle de Savage est un individu qui, dans ses choix d'actions, se comporte *comme si* il attribuait des probabilités et des niveaux d'utilité aux différentes conséquences de chaque action et *comme si*, parmi les actions disponibles, il choisissait celle qui procure la plus grande espérance mathématique d'utilité.<sup>11</sup>

Les exemples que nous venons de donner concernent un environnement statique dans lequel la relation entre le décideur et son environnement ne s'échelonne que sur une seule période: celle qui sépare la décision individuelle (le choix d'une action) de la réalisation de la conséquence de cette décision. On peut sans difficulté concevoir un type d'environnement plus complexe dans lequel l'acteur est en relation répétée avec son environnement, en provenance duquel il reçoit des signaux. De tels signaux peuvent alors être utilisés pour modifier les croyances probabilistes (subjectives dans le cadre savagien) qu'il forme sur les possibilités de réalisation des différentes conséquences de ses actions. Dans ce contexte, l'hypothèse de rationalité cognitive est usuellement interprétée comme impliquant un processus de révision par l'acteur de ces croyances probabilistes, en conformité avec la règle de Bayes.

<sup>10</sup>Voir Allais [1] pour une critique classique de ce modèle et Machina [38] pour un tour d'horizon pédagogique des approches alternatives.

<sup>11</sup>Il n'entre pas dans notre propos de rappeler les très nombreuses controverses qui ont entouré l'utilisation du critère savagien comme expression privilégiée de la rationalité. Ces problèmes sont examinés dans Picavet [45].

### 2.3 Un cas particulier de la rationalité cognitive: La rationalité axiologique

La discussion précédente portait sur des situations de choix en incertitude (probabilisable ou non) dans lesquelles la multiplicité des conséquences d'une action est liée au jeu de forces anonymes (mécanisme de sélection d'une boule dans une urne, distribution *a priori* des aptitudes d'un employé potentiel) dont il est difficile pour le décideur d'anticiper le résultat dans le cas d'espèce. Dans les termes de la discussion classique de Luce et Raiffa [37], ces situations appartiennent à la catégorie générale des *jeux contre la nature*. Mais il existe de nombreuses autres situations, dites de *jeux contre les personnes* ou, comme nous les appellerons par la suite, d'*interaction sociale*, dans lesquelles la multiplicité des conséquences d'une action individuelle résulte de décisions prises par d'autres individus. L'exemple simple suivant illustre ce type de situation.

Considérons deux cyclistes aux aptitudes sportives *a priori* égales qui doivent s'affronter dans une compétition internationale et qui doivent décider, avant celle-ci, de prendre ou non des stéroïdes anabolisants. Chaque cycliste est assuré d'une victoire s'il est le seul à consommer des stéroïdes mais la consommation de stéroïdes par les deux cyclistes n'augmente pas la probabilité de victoire de chacun. En outre, la consommation de stéroïdes affecte négativement la santé de chaque cycliste. Le problème de décision que doit résoudre un cycliste est schématiquement représenté par le tableau suivant.

	autre	cycliste
	stéroïdes	pas de stéroïdes
stéroïdes	<i>ex-aequo</i> en mauvaise santé	victoire en mauvaise santé
pas de stéroïdes	défaite en bonne santé	<i>ex-aequo</i> en bonne santé

Comme dans le cas d'un jeu contre la nature, il s'agit ici, du point de vue du cycliste, d'un problème de décision où plusieurs conséquences distinctes peuvent résulter du choix d'une seule action. Si le cycliste consomme des stéroïdes, il en résultera comme conséquence soit un *ex-aequo* en mauvaise santé, soit une victoire en mauvaise santé. De même, s'il s'abstient de consommer des stéroïdes, la conséquence sera soit un *ex-aequo* en bonne santé, soit une défaite en bonne santé. Pour cette raison, on peut, du point de vue du cycliste, aborder ce problème de décision avec les mêmes canons de rationalité cognitive que ceux usuellement déployés dans l'analyse des jeux contre la nature. De fait, supposons que chaque cycliste soit davantage préoccupé par sa performance sportive que par sa santé (au sens où il préfère une certitude de victoire en mauvaise santé à un *ex-aequo* en bonne santé et un *ex-aequo* en mauvaise santé à une certitude de défaite en bonne santé). Si notre cycliste obéit au modèle de Savage, et compare les deux actions disponibles en attribuant des probabilités aux différentes conséquences qu'elles peuvent avoir et en adoptant l'action qui lui donne en moyenne (pour les probabilités considérées) la meilleure conséquence, on peut prédire très précisément quelle action il choisira (même sans connaître ces probabilités). Il consommera des stéroïdes pour la simple raison que, quel que soit le choix de son adversaire, la conséquence de ce comportement est meilleure pour lui que celle du comportement alternatif. Le raisonnement qui mène à une telle conclusion est tout à fait analogue à celui qui est postulé dans le cas des jeux contre la nature. Il procède d'une rationalité instrumentale et cognitive du

cycliste tout à fait classique en ce sens qu'il prête au cycliste non seulement la poursuite d'un objectif particulier, mais également une certaine compréhension des conséquences de ses actions.

Mais la consommation de stéroïdes peut elle véritablement être dite "rationnelle" pour le cycliste ? Plus précisément le choix de cette action mène-t-il à la meilleure conséquence possible de son point de vue, et dans l'hypothèse d'une compréhension appropriée du problème de choix ? La réponse est évidemment négative *si l'on suppose qu'un comportement analogue est adopté par l'autre cycliste*. Or, précisément, la spécificité que présente une situation d'interaction sociale par rapport à un jeu contre la nature est que la multiplicité de conséquences que peut avoir une action ne résulte pas d'un système anonyme de forces (la nature) mais bien de décisions prises par des individus *a priori* analogues au décideur. Pour cette raison, l'opération, de nature *cognitive*, consistant à comprendre comment différentes conséquences peuvent résulter d'une action individuelle revient à comprendre comment d'autres individus semblables au décideur (du moins quant à leur implication dans l'interaction sociale considérée) prennent leur décision.

Si chacun des deux cyclistes consomme des stéroïdes, chacun se retrouvera dans une situation que chacun juge moins bonne que celle qui serait survenue s'ils s'étaient tous les deux abstenus d'absorber de telles substances. Comme dans toutes les situations de *dilemme du prisonnier*, dont cet exemple est une illustration, l'adoption par les individus d'un comportement rationnel (au sens instrumental et cognitif classique) entraîne des conséquences qui leur sont néfastes. Comme Amartya Sen [55] l'avait noté il y a plus de vingt ans, il n'est donc pas rationnel d'être rationnel. Pour le dire de manière moins provocante, la rationalité cognitive et instrumentale traditionnelle n'est pas nécessairement un bon guide de l'action individuelle dans les problèmes d'interaction sociale. Même s'il est "intéressé" (i.e. s'il se fonde exclusivement sur ce que lui-même juge bon ou bien), lorsqu'il examine si une certaine conduite de sa part est rationnelle, le participant à une interaction sociale doit se demander si cette conduite a de bonnes conséquences pour lui lorsque les autres participants adoptent des principes d'action analogues au sien. La raison qui rend nécessaire une telle hypothèse sur les motivations des autres participants est purement cognitive: l'acteur individuel n'a aucune raison spéciale de supposer des autres participants des principes d'action différents du sien.

Si l'on impose cette contrainte d'analogie des principes d'action, la rationalité cesse d'être une propriété de la seule action individuelle: elle devient la propriété d'une *norme sociale de comportements individuels*. Dans ces conditions, le choix rationnel de l'individu devient indissociable de l'adoption raisonnée d'une telle norme.

A partir de ce moment, on s'éloigne du noyau classique des conceptions instrumentales ou cognitives de la rationalité pour aborder ce que Raymond Boudon (voir [5], [6], [7]), dans le prolongement des travaux de Weber, a récemment désigné comme la conception "axiologique" de la rationalité. Mais la discussion précédente montre que la césure catégorielle entre le niveau "cognitif/instrumental" et le niveau "axiologique" peut disparaître dans certains cas. En effet, telle qu'elle est ici envisagée, la rationalité "axiologique" ne fait que préciser la rationalité cognitive, dans des situations d'interaction sociale dont le propre est de faire dépendre les conséquences des décisions individuelles des décisions d'autres individus. La connaissance qu'acquiert l'individu de son en-

vironnement doit alors, pour être rationnelle, reconnaître cette interdépendance des actions individuelles et exploiter la similitude des motifs. Comment reconnaître rationnellement cette interdépendance? Tel est le problème que pose la définition de la rationalité axiologique comme cas particulier de la rationalité cognitive. C'est pour ce problème que nous tentons maintenant de caractériser une solution définie, dans les cas où nous le pouvons.

### 3 Une théorie de la rationalité axiologique pour les problèmes d'interaction sociale à préférences données

#### 3.1 Présentation du cadre formel: Jeux et formes de jeux

Le point de départ de notre approche consiste à prêter à l'acteur individuel une capacité de compréhension de la nature même de l'interaction sociale. Cela suppose de l'individu une capacité à se représenter les éléments constitutifs fondamentaux de cette interaction tels qu'ils peuvent être appréhendés par la théorie des jeux moderne.

De manière générale, l'interaction sociale concerne les conséquences qui peuvent résulter de l'adoption, par les participants à l'interaction, de différents comportements. La description de l'interaction sociale doit donc comporter une description de chacun des éléments suivants.

1) **L'ensemble  $N = \{1, \dots, n\}$  des  $n$  individus impliqués dans l'interaction** (dans l'exemple des cyclistes, cet ensemble contient deux éléments).

2) Pour chaque individu  $i \in N$ , **l'ensemble  $A_i$  des actions  $a_i$  qui sont disponibles** (i.e. que l'individu  $i$  peut effectivement choisir). Dans l'exemple cycliste, ces actions étaient au nombre de deux: consommer des stéroïdes (action *ste*) ou s'en abstenir (action *abs*) mais on aurait pu décrire plus finement les comportements impliqués par ces deux actions en distinguant, par exemple, l'action de consommer un stéroïde anabolisant après le petit déjeuner de la journée précédant la course de la consommation de stéroïde après le diner de cette même journée. De fait, le concept d'"action" qui entre dans la description d'un problème d'interaction sociale est très général et peut désigner des règles complexes de décision, contingentes par rapport à la réalisation d'aléas extérieurs et/ou de choix d'actions par d'autres individus. Par exemple, l'action "ouvrir son parapluie quelle que soit la température observée" peut être considérée comme une action différente de celle consistant à ouvrir son parapluie en cas de pluie, de neige ou de grêle mais à le garder fermé dans tous les autres états du monde. De manière analogue, l'action consistant, pour un orateur terminant un exposé, à répondre poliment à toute question posée peut être distinguée de celle consistant à répondre de manière polie aux questions formulées de manière polie et à répondre de manière insolente aux questions formulées de manière insolente (si la situation d'interaction sociale supposée dans cet exemple permet de donner un sens clair aux critères de "politesse" et d'"insolence").

3) **L'ensemble  $\Omega$  des états sociaux** qui peuvent être atteints par les individus comme résultats (ou conséquences) du choix simultané de leurs actions. Un élément  $\omega \in \Omega$  s'interprète comme la description la plus fine qui soit, pour les besoins du problème considéré, de toutes les caractéristiques pertinentes de la

situation atteinte par les individus au terme de leur interaction. Ces caractéristiques pertinentes peuvent bien sûr porter sur les accomplissements effectifs des individus eu égard à certains critères objectifs (comme par exemple leur santé, leur revenu, la qualité de leur relation avec d'autres, etc). Elles peuvent également incorporer des éléments d'appréciation subjective que peuvent avoir les individus sur ces "accomplissements" (par exemple leur satisfaction). Dans l'exemple cycliste, l'ensemble  $\Omega$  comportait les quatre éléments suivants:

- *ex-aequo* en mauvaise santé pour les deux cyclistes ( $\omega_1$ ),
- victoire du premier cycliste en mauvaise santé et défaite du second en bonne santé ( $\omega_2$ ),
- *ex-aequo* en bonne santé pour les deux cyclistes ( $\omega_3$ ) et,
- défaite du premier cycliste en bonne santé et victoire du second en mauvaise santé ( $\omega_4$ ).

4) **Une fonction**  $r : \prod_{i \in N} A_i \rightarrow \Omega$  **de résultats** qui associe, à chaque configuration d'actions individuelles (une action par individu), un état social particulier. Pour une configuration d'actions  $(a_1, \dots, a_n) \in \prod_{i \in N} A_i$ , on interprète  $r(a_1, \dots, a_n)$  comme la *conséquence* de l'adoption simultanée, par les individus  $1, \dots, n$ , des actions  $a_1, \dots, a_n$ . Dans l'exemple cycliste,  $r(\cdot)$  est définie par:

$$r(\text{ste}, \text{ste}) = \omega_1, r(\text{ste}, \text{abs}) = \omega_2, r(\text{abs}, \text{ste}) = \omega_3 \text{ et } r(\text{abs}, \text{abs}) = \omega_4.$$

Nous serons quelquefois amenés à considérer les résultats de l'interaction du point de vue d'un individu générique, l'individu  $i$ . On notera alors  $r(a_i, \mathbf{a}_{-i})$  le résultat de l'interaction quand l'individu  $i$  choisit l'action  $a_i \in A_i$  et quand les individus autres que  $i$  choisissent la configuration d'actions  $\mathbf{a}_{-i} \in \prod_{j \neq i} A_j$ . Lorsque

nous envisagerons l'interaction du point de vue de deux individus génériques  $i$  et  $j$ , nous noterons  $(a_i, a_j, \mathbf{a}_{-ij})$  la configuration d'actions dans laquelle  $i$  choisit  $a_i$ ,  $j$  choisit  $a_j$  et les autres individus adoptent la configuration d'actions  $\mathbf{a}_{-ij} \in \prod_{h \neq i, j} A_h$ .

Pour l'étude abstraite de certains problèmes d'interaction sociale, la description contenue dans le quadruplet  $(N, \prod_{i \in N} A_i, \Omega, r(\cdot))$ , que l'on appelle, en théorie des jeux, une *forme de jeu*<sup>12</sup> suffit. Mais si l'on veut aboutir à des conclusions plus définies sur le résultat prévisible de l'interaction sociale, il est indispensable de compléter cette description, lorsqu'on le peut, par l'identification, pour chacun des  $n$  participants, d'un élément supplémentaire:

5) **un objectif individuel** ayant la forme minimale d'un critère  $R_i$  ( $i = 1, \dots, n$ ) de classement des états sociaux. Formellement, un tel critère  $R_i$  est une relation binaire sur  $\Omega$  et reçoit l'interprétation suivante: quels que soient les états sociaux  $\omega$  et  $\omega'$ ,  $\omega R_i \omega'$  signifie que  $\omega$  satisfait au moins autant l'intérêt de l'acteur  $i$  que  $\omega'$ . On désigne la préférence stricte de l'acteur par  $P_i$  ( $\omega P_i \omega'$  signifie que  $\omega$  satisfait strictement mieux l'intérêt de  $i$  que  $\omega'$ ) et l'indifférence par  $I_i$  ( $\omega I_i \omega'$  signifie que  $\omega$  et  $\omega'$  sont équivalents quant à leur capacité à satisfaire l'intérêt de  $i$ ). Comme il est d'usage, et pour les raisons qui ont été mentionnées plus haut, nous supposons que la relation binaire  $R_i$  est complète (elle permet de comparer n'importe quelle paire d'états sociaux distincts) et acyclique (elle n'engendre pas de cycles dans le classement strict des états

<sup>12</sup>Voir par exemple Deb [11], Deb, Pattanaik et Razzolini [12], Gärdenfors [22], Gaertner, Pattanaik and Suzumura [21], Moulin [39] et Peleg [43], [44]

sociaux qu'elle induit). Dans l'exemple des cyclistes, les objectifs des individus sont définis par les critères de classement suivants:  $\omega_2 P_1 \omega_4 P_1 \omega_1 P_1 \omega_3$  et  $\omega_3 P_2 \omega_4 P_2 \omega_1 P_2 \omega_2$

La donnée fournie par le quintuplet  $(N, \prod_{i \in N} A_i, \Omega, r(\cdot), \{R_i\}_{i=1}^n)$  forme ce que l'on appelle un *jeu* (sous forme normale). Dans de tels jeux, il peut arriver que deux (ou plusieurs) participants soient identiques du point de vue de leurs caractéristiques pertinentes pour l'interaction. Plus précisément, il peut arriver que deux individus aient le même ensemble d'actions disponibles et que, à une permutation de point de vue près, leurs préférences pour l'ensemble des conséquences de l'interaction soient les mêmes. De tels individus ne seront évidemment pas *physiquement* identiques. L'un pourra s'appeler Paul et avoir les yeux bleus tandis que l'autre se nommera Jacques et aura les yeux marrons. Ils n'auront pas non plus littéralement les mêmes préférences pour les conséquences de l'interaction sociale. Dans l'exemple précédent, le premier cycliste préfère un *ex-aequo* en bonne santé à une situation où il perd en bonne santé et où l'autre gagne en mauvaise santé, tandis que le second cycliste préfère la situation inverse. Leurs préférences s'opposent donc en ce sens. Mais elles sont identiques si l'on procède à un échange de rôle (ou de point de vue).

**Definition 1** *Les individus  $i$  et  $j$  sont identiques dans le jeu  $(N, \prod_{i \in N} A_i, \Omega, r(\cdot), \{R_i\}_{i=1}^n)$*

*si et seulement si  $A_i = A_j$  et  $\omega R_i \omega' \Leftrightarrow \widehat{\omega} R_j \widehat{\omega}'$  pour tous les états sociaux  $\omega, \omega', \widehat{\omega}$  et  $\widehat{\omega}'$  tels que  $\omega = r(a_1, \dots, a_i, \dots, a_j, \dots, a_n)$ ,  $\omega' = r(a'_1, \dots, a'_i, \dots, a'_j, \dots, a'_n)$ ,  $\widehat{\omega} = r(a_1, \dots, a_j, \dots, a_i, \dots, a_n)$  et  $\widehat{\omega}' = r(a'_1, \dots, a'_j, \dots, a'_i, \dots, a'_n)$  pour les configurations  $(a_1, \dots, a_i, \dots, a_j, \dots, a_n)$  et  $(a'_1, \dots, a'_i, \dots, a'_j, \dots, a'_n) \in \prod_{i \in N} A_i$ .*

Le problème est maintenant de proposer une notion de rationalité applicable aux comportements individuels dans les situations d'interaction sociale de cette classe.

### 3.2 De la rationalité cognitive unilatérale à la rationalité cognitive-axiologique multilatérale

Le choix d'une *stratégie dominante*, lorsqu'il en existe une, constitue un critère de rationalité cognitive particulièrement satisfaisant dans les situations d'interaction stratégique. Rappelons qu'une stratégie dominante pour un individu est une action qui a la propriété d'entraîner une conséquence meilleure que celle de n'importe quelle autre action disponible et ce, quelles que soient les actions adoptées par les autres participants. Dans le formalisme de la sous-section précédente, on définit cette notion comme suit.

**Definition 2** *Dans un jeu sous forme normale  $(N, \prod_{i \in N} A_i, \Omega, r(\cdot), \{R_i\}_{i=1}^n)$ ,*

*l'action  $\widehat{a}_i \in A_i$  est une stratégie dominante pour l'individu  $i$  si quel que soit  $\mathbf{a}_{-i} \in \prod_{j \neq i} A_j$ ,  $r(\widehat{a}_i, \mathbf{a}_{-i}) P_i r(a_i, \mathbf{a}_{-i})$  pour tout  $a_i \in A_i$ .*

Un individu peut, certes, ne pas avoir de stratégie dominante. Mais lorsqu'il en a une, le modèle usuel de la rationalité cognitive prévoit que l'individu l'adoptera. De fait, il a dans ce cas de 'bonnes' raisons de le faire. Une stratégie

dominante mène à la meilleure conséquence possible du point de vue de l'agent quoi que fassent les autres individus. Plus généralement, si tous les individus ont une stratégie dominante, la théorie de la rationalité cognitive traditionnelle prévoit qu'ils adopteront tous une telle stratégie. Mais les individus ont-ils raison d'adopter un tel comportement ? Ont-ils raison d'aborder l'interaction sociale en raisonnant comme si le comportement des autres était donné ?

Comme nous venons de le dire, ils ont des raisons individuelles de le faire. En adoptant une stratégie dominante, *chacun* fait mieux qu'en adoptant tout autre comportement quoi que fassent les autres. Mais la conséquence qui résulte de l'adoption par tous les individus de stratégies dominantes n'est pas nécessairement la meilleure possible du point de vue même des individus. Il n'est donc pas désirable pour les individus que tous (ou un grand nombre) adoptent ce principe d'action. Or un participant individuel à l'interaction n'a aucune raison spéciale de postuler une différence dans les motivations des uns et des autres. Pour cette raison, ce participant est conduit à comparer les conduites envisageables pour lui *sous la contrainte d'une motivation semblable des autres*.

En ce sens, dans les problèmes d'interaction sociale, la rationalité n'est plus seulement une propriété de l'action individuelle: elle devient la propriété d'une *norme sociale de comportements individuels* et le choix rationnel coïncide avec l'adoption d'une norme. Ainsi, dans un problème d'interaction sociale, une norme correcte de comportement individuel ne peut pas être caractérisée comme rationnelle seulement sur la base de l'évaluation individuelle des conséquences dans une perspective unilatérale sur l'action individuelle (i.e. sans exploiter la similitude prévisible des motifs des uns et des autres). Il faut également s'intéresser au comportement qu'adopteront les autres individus, et exploiter la similitude des motifs, qui découle de la similitude des situations individuelles.

Toute norme de comportement individuel doit donc être une norme *sociale* qui prescrit à chaque individu un comportement, et cette norme sociale peut faire l'objet d'une élucidation individuelle. Dans l'exemple des cyclistes, la norme qui prescrit à chaque individu l'adoption de sa stratégie dominante a, certes, la propriété de fournir aux uns et aux autres des raisons d'agir identiques. Il s'agit bien d'une norme sociale au sens de ce qui précède. Mais il en est d'autres. Et, comme l'exemple des cyclistes le révèle nettement, certaines peuvent conduire à de meilleures conséquences du point de vue des participants eux-mêmes. Il est alors rationnel de les comparer de ce point de vue.

### **3.2.1 Deux propriétés d'une norme rationnelle de comportements individuels**

Quelles pourraient être les propriétés d'une norme sociale rationnelle de comportements individuels, dans ce contexte ? Nous en apercevons deux. La première de ces propriétés est de prescrire aux individus des actions que tous les participants s'accordent à juger *analogues* du point de vue de la motivation de la conduite.

Le mot "analogue" est, certes, ambigu à ce stade de la discussion. Et de fait, nous n'entendons pas entièrement lever l'ambiguïté, parce qu'elle est liée aux différents systèmes de catégories que les agents peuvent employer lorsqu'ils décrivent et analysent leur situation. Des actions adoptées par des individus seront considérées comme analogue si elles peuvent être perçues par eux comme

étant de “même nature”, faisant alors l’objet d’un choix gouverné par des raisons semblables.

Ainsi, dans l’exemple cycliste, les actions de consommation de stéroïdes par chacun des deux cyclistes sont clairement considérées comme analogues (même si la dénomination “consommer des stéroïdes” peut renvoyer, chez les deux cyclistes, à des actions physiques distinctes: une consommation après le déjeuner pour l’un, après le dîner pour l’autre, etc.). Dans un cas comme celui-ci, à certaines variantes près, qui sont corrélées avec des caractéristiques individuelles non pertinentes pour le problème de choix considéré, les raisons de l’action sont les mêmes pour les différents agents.

Nous admettrons que la notion d’‘analogie’ entre actions est définie au moyen d’une liste ordonnée, et indicée par  $i$ , de  $n$  fonctions  $\Psi^i : A_i \rightarrow \prod_{j \in N} A_j$ , chacune associant à toute action de l’individu auquel elle correspond, l’action de chaque autre individu (y compris lui-même)<sup>13</sup> qui lui est “analogue”. Avec ce formalisme, écrire  $(a_1, \dots, a_n) = \Psi^i(a_i)$  signifie que les actions  $a_1$  de l’individu 1,  $a_2$  de l’individu 2, etc. sont ‘analogues’ à l’action  $a_i$  de l’individu  $i$ , au sens où leur choix est motivé par des *raisons semblables*. On notera  $\Psi_j^i(a_i)$  l’action de l’individu  $j$  qui, selon la notion d’analogie sous-jacente à la fonction  $\Psi^i$ , est ‘analogue’ à l’action  $a_i$  de l’individu  $i$ . Pour justifier l’interprétation d’une telle liste de fonctions en tant que *système d’analogies* entre actions, il convient de lui imposer quelques propriétés minimales, dont l’énoncé est l’objet de la définition suivante.

**Definition 3** *On dit d’une liste ordonnée de fonctions  $\Psi^i : A_i \rightarrow \prod_{j \in N} A_j$ , pour  $i \in N$ , qu’elle constitue un système d’analogies entre actions pour le jeu  $(N, \prod_{i \in N} A_i, \Omega, r(\cdot), \{R_i\}_{i=1}^n)$  si elle satisfait (i)  $\Psi_j^i(a_i) = a_j \Leftrightarrow \Psi_i^j(a_j) = a_i$  pour tous les individus  $i$  et  $j$  et pour toutes les actions  $a_i \in A_i$  et  $a_j \in A_j$ ; et (ii) pour toute paire d’individus  $i$  et  $j$  (non nécessairement distincts) identiques,  $\Psi_j^i(a_i) = a_i$  pour toute action  $a_i \in A_i$ .*

La propriété (i) n’est rien d’autre que la traduction formelle de l’intuition banale suivant laquelle la relation d’analogie entre objets, telle qu’elle est habituellement conçue, est symétrique. Si un objet  $a$  est analogue à un objet  $b$ , alors l’objet  $b$  doit aussi être analogue à l’objet  $a$ . La propriété (ii) ne fait que traduire formellement le principe leibnizien de l’identité des indiscernables. Elle requiert en effet de la notion d’*analogie* sous-jacente à la liste de fonctions  $(\Psi^1, \dots, \Psi^n)$  qu’elle coïncide avec celle d’*identité* lorsqu’elle s’applique à des individus identiques au sens de la définition 1. En clair, pour des individus dont les caractéristiques pertinentes pour l’interaction sociale sont identiques, les actions sont analogues (motivées par des raisons semblables) si, et seulement si, elles sont identiques. On comprendrait mal en effet qu’une notion de similitude conduite à considérer comme *analogues* des actions *distinctes* d’individus qui, du point de vue de leurs caractéristiques pertinentes pour l’interaction, sont *identiques*.

Si les deux propriétés explicitement mentionnées dans la définition 3 paraissent s’appliquer assez naturellement à une théorie de l’analogie ou de la similitude, il n’en va peut être pas de même de l’exigence implicite que nous imposons

<sup>13</sup>Le fait que  $\Psi_i$  admette comme image le produit cartésien de tous les ensembles de stratégies (plutôt que le produit cartésien des ensembles de stratégies des individus autres que  $i$ ) n’est qu’une commodité de notation qui allège l’écriture des propriétés 1 et 2 ci-dessous.

à une telle théorie: prendre la forme d'un système de *fonctions*. De fait, nous imposons par cette hypothèse qu'à une action particulière d'un individu ne corresponde qu'une seule action d'un autre individu qui lui soit équivalente. En particulier; deux actions d'un même individu ne peuvent pas être équivalentes entre elles. Nous ne pouvons à ce stade donner de meilleure justification de cette hypothèse que celle de la facilité technique. Mais une théorie plus complète devrait certainement envisager l'assouplissement de cette hypothèse.

Admettons donc qu'il existe, pour les participants à l'interaction, un système d'analogies entre actions qui soit reconnu par eux. L'origine et la nature exacte de ce système ne nous intéressent pas ici. Nous admettons qu'un tel système existe, et qu'il vérifie les propriétés de la définition 3. La première propriété de rationalité d'une norme sociale  $\nu = (\nu_1, \dots, \nu_n) \in \prod_{i \in N} A_i$  de comportements individuels est alors la suivante:

*Analogie des actions individuelles:*  $\nu = \Psi^i(\nu_i)$  pour tout  $i \in N$ .

Le fait qu'une norme sociale de comportements individuels prescrive aux individus des actions analogues (au sens du système d'analogies  $(\Psi^1, \dots, \Psi^n)$ ) paraît être une exigence minimale de rationalité dans le contexte qui est le nôtre ici. Comment un individu pourrait-il admettre que le principe qui l'amène à choisir son action soit différent du principe qui gouverne le choix effectué par les autres ? Dans la mesure où l'individu qui participe à une interaction sociale comprend que le résultat de sa décision dépend de choix effectués par des individus qui sont confrontés à un problème de décision analogue au sien, il doit formuler des hypothèses sur les *raisons* qui amèneront ces autres individus à faire ces choix. Dans la mesure où l'individu comprend l'analogie de situation entre ces autres individus et lui-même, l'hypothèse suivant laquelle les autres individus vont agir selon une catégorie de motivation analogue à la sienne est la seule qui soit tenable d'un point de vue cognitif.

La seconde propriété qu'il paraît naturel d'imposer à une norme  $\nu = (\nu_1, \dots, \nu_n) \in \prod_{i \in N} A_i$  de comportement est de mener à un résultat que chacun juge meilleur (au moins faiblement) que n'importe quelle autre conséquence qui pourrait résulter de l'adoption, par les individus, de toute autre configuration d'actions analogues. Formellement, nous exigeons en effet d'une norme la propriété suivante:

*Rationalité individuelle :* Pour tout  $i \in N$ ,  $r(\Psi^i(\nu_i)) R_i r(\Psi^i(a_i))$  pour tout  $a_i \in A_i$

Une norme dont l'adoption ne produirait pas la meilleure conséquence possible, du point de vue des individus eux-mêmes, lorsque les acteurs agissent de manière analogue, ne saurait prétendre au statut d'une norme rationnelle de comportement. En effet, les préférences étant données (d'une manière qui synthétise l'ensemble des valeurs et des intérêts jugés pertinents par les individus), quelles raisons les individus pourraient-ils avoir de reconnaître une norme de comportements individuels analogues qui mène à de moins bonnes conséquences, pour chacun d'eux, qu'une autre norme respectant elle aussi l'analogie des prescriptions? Ces raisons ne pourraient être tirées ni de ce qu'ils valorisent, ni de ce que valorisent les autres. Il est alors plausible d'affirmer que de telles raisons sont inexistantes.

Ces deux propriétés de rationalité des normes individuelles de comportement pour les situations d'interaction sociale à préférences données ont été proposées dans Bilodeau et Gravel [?] dans un autre contexte. Si elles paraissent constituer une exigence de rationalité plausible pour une norme sociale de comportements individuels, il faut bien comprendre que cette exigence est forte. Il n'est pas difficile (voir Bilodeau et Gravel [?]) de présenter des exemples de jeux qui n'admettent *aucune* norme de comportement rationnel au sens de ces deux propriétés, et ce, quel que soit le système d'analogie auquel on puisse penser. Mais l'absence de normes axiologiquement rationnelles de comportements individuels dans certains problèmes d'interaction sociale ne doit pas nous surprendre davantage que, par exemple, l'absence (très fréquente) d'équilibres en stratégie dominante. Les situations d'interaction sociale sont des situations complexes. On ne peut donc pas s'attendre à ce que l'analyse de ces situations révèle toujours une manière correcte de se comporter capable de s'imposer avec évidence.

### 3.2.2 La rationalité axiologique: des prescriptions claires dans les jeux symétriques

Les prescriptions particulières de comportement auxquelles peuvent conduire les deux propriétés de rationalité que nous venons de mentionner, lorsque celles-ci sont mutuellement compatibles, dépendent à l'évidence de la conception de l'analogie sur laquelle elles sont fondées. Mais il est une classe de problèmes d'interaction sociale pour lesquels ces deux propriétés conduisent à des prescriptions claires: la classe des jeux symétriques. Rappelons qu'un jeu est symétrique si tous les participants à ce jeu sont identiques au sens de la définition 1. Dans des jeux symétriques, il ne peut y avoir qu'un seul système d'analogie satisfaisant la propriété 1) de la définition 3), à savoir, celui que définissent les fonctions identités. Les actions de deux individus sont analogues si et seulement si elles sont identiques. L'exemple des cyclistes mentionné plus haut décrit une situation de ce type. Dans une telle situation, la rationalité d'une norme s'énonce simplement. Est rationnel tout comportement qui, sous la contrainte de l'adoption par tous, conduit à la meilleure conséquence possible pour l'individu.

Les trois premiers exemples utilisés en introduction constituent de bonnes illustrations de ce type de jeu. Dans l'exemple cycliste, le refus de la consommation de produit dopants s'impose naturellement comme une norme correcte de comportement puisque, sous l'hypothèse de l'adoption par tous, il conduit à une meilleure conséquence que le dopage de tous.

Il en va de même des utilisateurs du lieu public qui peuvent contribuer à la propreté du lieu en s'abstenant d'y jeter leurs détritrus, mais qui ont une incitation personnelle à les jeter. A niveau donné de contribution personnelle à la propreté du parc, chaque promeneur préfère un parc propre à un parc qui l'est moins mais à niveau donné de contribution des autres à la propreté du parc, chaque promeneur préfère jeter ses détritrus. Ces utilisateurs ont donc des préférences sur les conséquences de leur interaction (le niveau de propreté du lieu et la combinaison de leurs contributions personnelles) qui, à une permutation de point de vue près, sont identiques. Et ils font face à des ensembles de stratégies (jeter ses détritrus et ne pas les jeter) qui sont identiques.

Le comportement de vote se laisse également représenter sous la forme d'un jeu symétrique. Ainsi, l'argument "il faut voter, car si de trop nombreuses personnes s'abstenaient de prendre la peine d'aller voter, la démocratie ne fonc-

tionnerait pas bien”, qui soutient l’unique norme de comportement vérifiant nos deux propriétés, est d’ordinaire jugé convaincant, et enseigné dans les écoles. Il est convaincant parce que, tout d’abord, nous considérons comme des données certains éléments relatifs aux préférences: la démocratie est une bonne chose (pour tous, elle est préférable à l’absence de démocratie). Sur cette base, nous estimons que le bon fonctionnement de la démocratie *avec participation individuelle* constitue, pour chacun, un avantage par rapport à une situation “sans démocratie”. Il est évidemment possible que chacun préfère encore la situation où la démocratie fonctionnerait sans sa participation, mais l’absence de participation ne peut, sans ruiner la démocratie (donc sans violation de la condition de rationalité individuelle) être l’objet d’une injonction normative suivie par tous.

Il est cependant un grand nombre de situations qui ne se laissent pas aisément représenter sous la forme de jeux symétriques et pour lesquelles, par conséquent, les deux propriétés de rationalité des normes identifiées dans la sous-section précédente sont moins concluantes. Considérons par exemple le cas d’individus qui doivent contribuer financièrement à un bien public. Un exemple d’une telle situation est fournie par l’utilisation des transports publics parisiens qui, d’ailleurs, donnent lieu à de fréquents comportements de passager clandestin. Certes, beaucoup d’usagers franchissent les portes d’accès aux quais du métropolitain munis de titres de transport valides, même si la sévérité et la fréquence des sanctions prévues en cas de resquillage ne suffisent probablement pas à les inciter, à elles seules, à cette conduite vertueuse. On pourrait penser qu’en agissant ainsi, ils appliquent une norme de non-resquillage en se tenant le raisonnement suivant lequel “si chacun resquillait, il n’y aurait plus de transport collectif” (raisonnement repris dans certaines campagnes d’affichage destinées à lutter contre le comportement de passager clandestin). Mais la norme suivant laquelle chacun devrait payer pour un ticket de la régie des transports à un prix exigé par cet organisme satisfait-elle nos deux propriétés, relativement à un système plausible d’analogies ? Et quelle forme un tel système prendrait-il ?

En tout état de cause, un tel système n’aurait aucune raison d’être fondé sur la fonction identité. Les contributeurs potentiels au bien public “transports en commun” que sont les usagers de ces transports ne sont pas, en réalité, des individus qui doivent se considérer comme identiques les uns aux autres. Ils n’ont peut être pas les mêmes préférences pour les différentes combinaisons de la qualité du service qui leur est offert et de leur contribution personnelle. Mais surtout, ils n’ont certainement pas la même capacité contributive, c’est à dire, dans le langage formel de la sous-section précédente, le même ensemble de stratégies. Pour cette raison, il est peu vraisemblable que les usagers des services publics parisiens admettent que deux contributions sont analogues si, et seulement si, elles sont identiques. Il est peu vraisemblable qu’une personne qui gagne 2500F par mois considère que la contribution qu’on exige d’elle représente le même effort contributif que la même contribution exigée d’un cadre supérieur dont le revenu mensuel est de 100 000F. Et de fait, on peut penser qu’un certain nombre d’individus qui empruntent les transports publics sans payer, ou qui ne font pas montre du respect le plus scrupuleux des lois fiscales au moment de remplir leur déclaration d’impôts, parviennent à légitimer leur comportement frauduleux en considérant que la contribution qui est légalement exigée d’eux n’est pas la bonne, parce qu’elle ne respecte pas le type de symétrie qu’ils jugent pertinent. Contrairement peut être au cycliste qui s’injecte des stéroïdes, les resquilleurs des transports publics parisiens n’ont pas toujours l’impression

de violer une norme de comportement qu'ils jugent eux-mêmes rationnelle.

Mais sur quel système d'analogies doit-on fonder une norme rationnelle de comportements individuels lorsque les individus ne sont pas identiques au sens de la définition 1? *A priori*, des principes très variés pourraient faire l'affaire. Ainsi, on pourrait considérer deux contributions financières comme analogues si elles représentent la même fraction de la richesse des individus, ou bien encore, si elles représentent le même sacrifice psychologique (qui devrait alors être soigneusement défini), etc. Mais il sera alors difficile, en fait, de garantir l'existence de normes vérifiant les deux propriétés de rationalité avancées plus haut. Par exemple, il n'est pas vrai en général qu'une norme qui prescrit à tous les usagers d'un service public de contribuer à la production de ce service au niveau que chacun préfère sous l'hypothèse que tous les usagers contribuent à hauteur de la même fraction de leur richesse, prescrive de fait à ces usagers de contribuer des fractions égales de leur richesse. Dans les termes de la théorie exposée, il y a alors contradiction entre les exigences de la *rationalité individuelle* et de *l'analogie des actions*.<sup>14</sup> Dès que l'on quitte le terrain des jeux symétriques, il devient difficile d'identifier des normes axiologiquement rationnelles de comportement individuel. L'enquête peut alors se donner pour mission d'identifier, dans les arguments qui servent à "se faire la morale" les uns aux autres, des efforts pour décrire la situation d'une manière qui la ramène, implicitement, à un jeu symétrique. On peut aussi essayer de comprendre les éléments de dissymétrie jugés pertinents et importants par les individus, et qui les empêchent, précisément, d'accepter de voir dans la situation concernée un jeu symétrique.

## 4 Analyse des droits:Quelques pistes

### 4.1 Liberté de choix et préférences

Dans de très nombreux cas, l'adoption de normes est liée à un effort de compréhension de situations qui sont appréhendées d'une manière très peu définie. Par exemple, il est rare que l'on mette en cause l'importance politique du principe de la liberté de culte (que ce soit d'ailleurs pour l'approuver ou pour le condamner). Or, ce principe vaut pour une interaction sociale générique, définie à partir d'éléments en eux-mêmes peu informatifs: il y a plusieurs personnes; elles coexistent; elles peuvent s'engager (seules ou à plusieurs) dans des activités religieuses, et aussi dans des activités qui causent des désagréments, voire un empêchement, aux personnes qui pratiquent des activités religieuses; les actions des uns et des autres peuvent être soumises à des restrictions normatives. En l'absence d'information supplémentaire sur les intentions ou les préférences des agents, rien ne permet d'apercevoir une issue prévisible aux interactions concernées.

Il semble pourtant essentiel, pour certaines normes, que l'on puisse les spécifier et les justifier en dehors de toute hypothèse spécifique sur les préférences. Ainsi, les normes constitutionnelles font abstraction des forces et des aspirations politiques en présence, les droits de l'homme sont "universels"; les règles morales sont "absolues", etc. Est-il possible de caractériser des éléments de rationalité axiologique dans des contextes aussi peu définis?

---

<sup>14</sup>Voir Bilodeau et Gravel [?] pour une discussion plus approfondie des problèmes de contributions financières à des biens publics dans ce contexte.

Envisagé dans les termes du modèle de la section précédente, le problème de justification que posent de telles normes renvoie au problème du choix d'une *procédure* (que l'on peut assimiler à une forme de jeu) par des individus qui ne connaissent pas les préférences qui les animeront lorsque cette forme de jeu encadrera leur interaction. Les exemples concrets de ce type de choix sont nombreux. Par exemple, les membres du jury d'un examen oral doivent convenir à l'avance d'une procédure d'audition et d'évaluation des candidats, ainsi que de la procédure à suivre au moment des délibérations, sans connaître ces candidats (ni donc, *a fortiori*, leurs préférences subjectives pour ceux-ci). Un second exemple, emprunté à Kreps [35], est celui de la décision d'aller au restaurant. En choisissant un restaurant, un individu choisit un lieu dans lequel il sera ultérieurement amené à faire un autre choix: celui d'un menu ou d'un plat. Typiquement, l'individu qui choisit un restaurant ne connaît pas précisément les préférences qui gouverneront son choix d'un menu particulier.

On peut de la même manière concevoir un système de droits individuels (par exemple les droits fondamentaux des personnes) comme le résultat d'une décision prise par des individus, en vue d'édicter par avance des règles qui devront gouverner l'interaction sociale, et cela, dans l'ignorance des objectifs qu'eux mêmes ou d'autres poursuivront dans cette interaction.

Les propriétés rationnelles de telles procédures ne peuvent évidemment être établies sur la base des objectifs particuliers visés par les individus. Elles doivent plutôt être définies en faisant référence à un ensemble *a priori* assez large de préférences que les individus *pourraient* avoir. Cette ignorance des individus quant aux préférences qui guideront leur comportement dans l'interaction sociale conduit naturellement à privilégier des critères de rationalité procédurale fondés sur la *flexibilité* et la *liberté* que les procédures donnent aux individus. Comme l'a montré Kreps [35], il existe en effet un lien très étroit entre le désir que peut avoir un agent individuel de disposer de flexibilité ou de liberté et son ignorance des préférences qui guideront ses choix dans le futur.

De fait, dans le choix d'un restaurant, comme dans celui d'un système de droits individuels, le choix qui est effectué est celui d'un *ensemble d'opportunités* dans lequel l'individu choisira ultérieurement une *option*. Dans le cas du restaurant, l'option correspond à un menu ou à un plat. Elle est pratiquement indistincte de la conséquence que le choix final de cette option entraîne pour le décideur, à savoir, la dégustation d'un certain repas. Un critère intéressant de rationalité dans le choix d'un ensemble d'opportunités dont les options sont indistinctes des conséquences qu'elles peuvent avoir est celui, caractérisé par Kreps [35], de la préférence pour la flexibilité ou la liberté de choix, qui a été étudiée dans une série de travaux récents.<sup>15</sup>

## 4.2 Le problème général du choix d'une procédure

Rationaliser les normes prenant la forme de droits reviendrait donc à formuler des critères de classement des "menus d'options", d'une manière pertinente pour certains problèmes d'interaction sociale. Or ces problèmes d'interaction sociale introduisent un degré de complexité supplémentaire, par rapport au problème du choix d'un restaurant, considéré par Kreps [35] et par les travaux postérieurs

<sup>15</sup> Voir par exemple Arrow [2], Bossert, Pattanaik et Xu [?], Bossert [?], Gravel ([25],[26]), Jones et Sugden [29], Klemisch-Ahlert [33], Pattanaik et Xu [42], Puppe ([47], [48]) et Sen ([56], [57], [58]).

sur la liberté de choix. Il faut en effet distinguer les options auxquelles ont accès les individus des conséquences que leur choix peut avoir. Les conséquences, en effet, dépendent de l'action des autres. Mais si l'on fait abstraction de cette difficulté supplémentaire, pour le moment non résolue, le choix d'un système de droits - ou d'une procédure - se poserait formellement comme celui d'une forme de jeu assignant à chaque participant l'ensemble des actions qu'il peut choisir et le résultat qu'il peut attendre de chaque configuration de comportements admissibles des uns et des autres.

Formellement donc, une procédure appliquée à un problème d'interaction sociale  $S = (N, \prod_{i \in N} A_i, r(\cdot))$  est un sous-ensemble  $\nu$  de  $\prod_{i \in N} A_i$ . La  $i$  ème composante de  $\nu$  s'interprète comme l'ensemble des actions qui sont *normativement légitimes* pour l'individu  $i$  dans la procédure considérée. Notons que, dans cette perspective,  $\prod_{i \in N} A_i$  (l'ensemble de toutes les configurations physiquement envisageable d'actions individuelles) est lui-même une procédure d'interaction sociale. Il s'agit d'une procédure que l'on peut appeler *état de nature*, et dans laquelle, pour chaque individu, est légitime tout ce qui est physiquement possible.

A notre avis, toute entreprise visant à justifier un critère particulier pour le choix d'une forme de jeu devrait prendre en considération cet "état de nature". En effet, l'état de nature décrit l'interaction sociale en l'absence de toute procédure d'encadrement normatif. A ce titre, il sert de point de référence pour l'évaluation des systèmes de normes possibles, car c'est en postulant l'absence de contraintes que l'on peut procéder à *l'arbitrage le plus ouvert possible* entre les dispositifs envisageables, en examinant ce qu'ils apportent à ceux qui participent à l'interaction sociale. L'importance de la référence à un état de nature (non contraint ou contraint seulement par la "loi naturelle") dans les théories politiques classiques est bien connue. Pourtant, et si l'on excepte le récent travail de Fleurbaey et Van Hees [?], ce concept n'a guère été utilisé jusqu'à présent dans les approches en termes de "formes de jeu" des systèmes normatifs.<sup>16</sup>

Par comparaison avec un état de nature initial, les dispositifs normatifs apportent aux individus des garanties plus fines que celles qu'ils possédaient. Par exemple, sous un dispositif garantissant la liberté de culte, les individus qui pratiquent un culte sont assurés de parvenir, s'ils le souhaitent et s'ils agissent en conséquence, à un état du monde dans lequel ils pratiquent ce culte sans rencontrer d'empêchement (alors que dans l'état de nature ils sont seulement certains de pouvoir pratiquer avec ou sans empêchement). Ce rétrécissement de l'ensemble des conséquences qui peuvent résulter d'une certaine action est obtenu grâce à des restrictions sur les possibilités d'action d'autrui.

On voit donc se dessiner un *jeu constitutionnel* dont l'enjeu est un *arbitrage* entre les *demandes de garanties* de certains individus et les interdictions qui résulteraient, pour les autres, de la mise en place de ces garanties. Si l'on admet que la contrainte ne se justifie ni par elle-même, ni par une volonté transcendante, seules les garanties offertes aux uns par les contraintes imposées aux autres sont capables de justifier celles-ci.

On reste ici dans une perspective conséquentialiste. Ce sont les garanties relatives aux conséquences de l'interaction sociale qui justifient les contraintes imposées aux participants. Mais comment s'entendre, précisément, sur les critères

<sup>16</sup> Voir par exemple Deb [11], Deb, Pattanaik et Razzolini [12], Gardenförs [22], Gaertner, Pattanaik et Suzumura [21] et Peleg [44].

d'arbitrage?

Une stratégie de réponse consisterait à postuler que chacun a la capacité d'apercevoir un accord possible fondé sur des “bonnes raisons” individuelles identiques. En effet, tous les participants peuvent s'efforcer de comprendre le fonctionnement de l'interaction sociale, et la manière dont un système de droits pourrait l'encadrer. Cette hypothèse sous-tend l'approche “politique” des droits esquissée dans Picavet [46]. Les agents se reconnaissent les uns aux autres une aptitude à la compréhension. En tant que “citoyens” potentiels, ils peuvent admettre que tous ont des positions semblables, à un certain niveau de description, dans l'interaction sociale. Sur cette base, chacun se demande s'il peut accepter les principes constitutionnels pour des *raisons identiques* à celles qui conduisent les autres à les accepter. Le recours à cette hypothèse de cognition suppose un effort pour mettre en évidence les “raisons” de considérer comme similaires des situations individuelles qui resteront malgré tout distinctes. En particulier, il faut pouvoir admettre que tous possèdent, après la concrétisation des garanties visées, des facultés de contrôle substantiellement similaires. La démarche proposée revient ainsi à tirer les conséquences des contraintes formelles sur l'argumentation ou la formation des jugements qui constituent, selon Kalleberg et Preston [31] le noyau central des approches cognitivistes de l'éthique: *impartialité* (égale attention aux différentes personnes); *cohérence* éthique ou *universalité* (traitement similaire des cas similaires, application uniforme des jugements); *compréhension* des faits pertinents (aussi complète et précise que possible).

Un tel schéma d'explication se situerait dans le prolongement direct du modèle proposé dans la section précédente. Le rôle de la “norme sociale d'actions individuelles” est tenu par ce que l'on pourrait appeler une “*norme sociale de revendications individuelles*”, qui borne les prétentions individuelles en les limitant à ce qui est compatible avec l'existence de “bonnes raisons” d'acceptation identiques pour tous.

## 5 Conclusion

Nous n'avons postulé aucune opposition de principe entre rationalité axiologique et conséquentialisme, puisque nous avons donné le premier rôle aux résultats de l'interaction sociale et aux garanties relatives à ces résultats. C'est dans cette perspective que nous avons tenté d'expliquer la sélection de certaines normes permettant d'agir “par principe”. Mais la sélection des normes elles-mêmes repose sur la manière dont elles garantissent ou excluent, pour les agents, l'accès à certains ensembles d'états sociaux.

L'analyse proposée pour l'exemple des cyclistes mettait en relief le fait que, dans les problèmes d'interaction sociale, la rationalité correctement comprise n'est plus la propriété exclusive d'une action individuelle, mais caractérise une *norme sociale de comportements individuels*. Cela n'implique aucune rupture avec la méthodologie individualiste, puisque c'est nécessairement le point de vue individuel des acteurs qui permet de formuler des propriétés de rationalité pour ces normes. Dans le langage courant, cette dimension multilatérale de la rationalité se traduit par des formulations abstraites et générales qui peuvent déconcerter: “il ne faut pas mentir”, “il faut respecter ses engagements contractuels”, etc. L'aspect inconditionnel (“il faut...”) et l'aspect distributif

(“chacun doit...”) peuvent paraître absurdes car il peut toujours y avoir de bonnes raisons (tirées par exemple de la conduite observée chez les autres) de ne pas faire ce qui est prescrit. Pourtant, on peut en rendre compte si l’on considère qu’il s’agit de prescriptions individuelles associées à des normes sociales rationnelles (rationalisables pour et par chaque individu) pour une classe donnée de situations d’interaction sociale.

L’exemple des cyclistes permettait de mettre en évidence l’existence de bonnes raisons pour adopter un autre comportement que celui que prescrit l’équilibre en stratégies dominantes. Il y aurait quelque irrationalité, pour chacun, à se déterminer sans exploiter la connaissance qu’il peut avoir de la similitude entre les motifs des uns et des autres. Cette analyse se laisse généraliser par une stratégie d’explication plus abstraite. Il ne s’agit d’ailleurs pas d’un schéma universel d’explication des normes: dans certains cas, il ne donnera aucune conclusion définie. Mais nous pensons que ce schéma est utile pour comprendre et synthétiser les “bonnes raisons” saillantes qui apparaissent dans un certain nombre d’exemples classiques de la rationalité en valeur considérée comme distincte de la promotion instrumentale de l’intérêt (participation volontaire au vote ou à la production d’un bien public; abstention “par principe” de la tricherie, du mensonge, du plagiat, de la corruption et de la fausse promesse; soumission volontaire à une norme contraignante mais favorable au “bien commun”, etc.).

Système de similitude sous-jacent et reconnaissance mutuelle de l’égale rationalité (au sens de l’aptitude à se déterminer par des raisons comparables): il nous reste à retrouver, dans le cas des choix constitutionnels, sous une forme différente, les éléments utilisés dans les cas plus simples où les préférences des individus sont connues et utilisées pour définir les normes. Dans ce contexte, le modèle suggère de s’intéresser à des systèmes normatifs par lesquels certaines demandes de garanties sont satisfaites, éventuellement au détriment de certaines autres. On pourrait coordonner de la sorte, par des raisons tirées de la situation des agents dans l’interaction, les explications de l’*acceptation* de différentes normes de liberté publique. Selon ce schéma, les normes apparaissent comme l’objet de prétentions légitimes, dans la mesure où, en matière de garanties et de restrictions dans le contrôle de la réalisation des états sociaux, elles ne représentent rien, pour personne, qui ne soit aussi attribué ou imposé à tout autre de manière analogue. Mais le schéma explicatif des normes qui prennent la forme de restrictions sur les ensembles d’actions des individus dans des cas où ces derniers ne sont pas encore déterminés quant à leurs préférences reste à préciser.

## References

- [1] M. Allais. Le comportement de l’homme rationnel devant le risque: Critique des postulats de l’école américaine. *Econometrica*, 21:503–546, 1953.
- [2] K. J. Arrow. A note on freedom and flexibility. In K. Basu, P. Pattanaik, and K. Suzumura, editors, *Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen*, chapter 1, pages 7–15. Oxford University Press, Oxford, 1995.
- [3] D. Bernheim. Rationalizable strategic behavior. *Econometrica*, 52:1007–1028, 1984.

- [4] M. Bordignon. Was kant right? voluntary provision of public goods under the principle of unconditionnal commitment. *Economic Notes*, 3:342–372, 1990.
- [5] R. Boudon. *Le Juste et le Vrai*. Fayard, Paris, 1995.
- [6] R. Boudon. Au delà du modèle du choix rationnel. In B. Saint-Sernin, E. Picavet, R. Fillieule, and P. Demeuenaere, editors, *Les Modèles de L'action*. Presses Universitaires de France, Paris, 1998.
- [7] R. Boudon. La rationalité axiologique. In S. Mesure, editor, *La Rationalité Des Valeurs*, pages 15–57. Presses Universitaires de France, Paris, 1998. collection "Sociologies".
- [8] R. Boudon. *Le sens des valeurs*. Presses universitaires de France, Paris, 1999.
- [9] R. Boudon, A. Bouvier, and F. Chazel. *Cognition et sciences sociales*. Presses universitaires de France, Paris, 1997.
- [10] J. S. Coleman. *Foundations of Social Theory*. Harvard University Press, Cambridge, MA, 1990.
- [11] R. Deb. Waiver, effectivity and rights as game forms. *Economica*, 61:167–178, 1994.
- [12] R. Deb, P. Pattanaik, and L. Razzolini. Game forms, rights and the efficiency of social outcomes. *Journal of Economic Theory*, 72:74, 1997.
- [13] P. Demeulenaere. *Homo oeconomicus*. Presses universitaires de France, Paris, 1996.
- [14] P. Demeulenaere. La légitimation et la dénonciation de la recherche de l'argent dans la modernité. *Archives de philosophie du droit*, 42:137–151, 1997.
- [15] P. Demeulenaere. Les ambiguïtés constitutives du modèle du choix rationnel. In B. Saint-Sernin, E. Picavet, R. Fillieule, and P. Demeulenaere, editors, *Les Modèles de L'action*. Presses universitaires de France, Paris, 1998.
- [16] A. Etzioni. Toward a kantian socio-economics. *Review of Social Economy*, 45:37–47, 1987.
- [17] J. Ferejohn and M. P. Fiorina. The paradox of not voting: A decision theoretic analysis. *The American Political Science Review*, 68:525–536, 1974.
- [18] R. Fillieule. La question de la rationalité dans les sciences sociales. Thèse de doctorat, université de Paris IV, 1994.
- [19] R. Fillieule. Frames, inferences and rationality: Some light on the controversies about rationality. *Rationality and Society*, 8, 1996.

- [20] R. Fillieule. Trois questions fondamentales pour les sciences sociales. In R. Fillieule and B. Valade, editors, *Introduction Aux Sciences Sociales*. Presses universitaires de France, Paris, 1996.
- [21] W. Gaerner, P. Pattanaik, and K. Suzumura. Individual rights revisited. *Economica*, 59:161–177, 1992.
- [22] P. Gärdenfors. Rights, games and social choice. *Nous*, 15:341–356, 1981.
- [23] A. Giddens. *La constitution de la société*. Presses universitaires de France, Paris, 1987. Traduction française de *The Constitution of Society* (Cambridge, Polity Press, 1984).
- [24] J. Goldthorpe. Rational action theory for sociology. *British Journal of Sociology*, 49:167–182, 1998.
- [25] N. Gravel. Can a ranking of opportunity sets attach intrinsic importance to freedom of choice? *American Economic Review*, 84:454–458, 1994.
- [26] N. Gravel. Ranking opportunity sets on the basis of their freedom of choice and their ability to satisfy preferences: A difficulty. *Social Choice and Welfare*, 15:371–382, 1998.
- [27] D. P. Green and I. Shapiro. Choix rationnels et politique: Pourquoi en savons-nous toujours aussi peu ? *Revue française de science politique*, 45:96–130, 1995.
- [28] J. Guttman. Understanding collective action: Matching behavior. *American Economic Review: Papers and Proceedings*, 68:346–358, 1978.
- [29] P. Jones and R. Sugden. Evaluating choices. *International Journal of Law and Economics*, 2:47–65, 1982.
- [30] D. Kahneman and A. Tversky. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5:207–232, 1974.
- [31] A. Kalleberg and L. Preston. Normative analysis and the problem of justification: The cognitive status of basic political norms. *The Journal of Politics*, 37:650–684, 1975.
- [32] E. Kant. *Fondements de la métaphysique des moeurs*. Delagrave, Paris, 1967. traduction de V. Delbos.
- [33] M. Klemisch-Ahlert. Freedom of choice: A comparison of different rankings of opportunity sets. *Social Choice and Welfare*, 10:189–207, 1993.
- [34] S. C. Kolm. *La Bonne Economie: La Réciprocité Générale*. Presses universitaires de France, Paris, 1984.
- [35] D. M. Kreps. A representation theorem for 'preference for flexibility'. *Econometrica*, 47:565–577, 1979.
- [36] J. J. Laffont. Macroeconomic constraints, economic efficiency and ethics: an introduction to kantian economics. *Economica*, 42:430–437, 1975.

- [37] R. D. Luce and H. Raiffa. *Games and Decisions*. John Wiley, New York, 1957.
- [38] M. Machina. *The Economic Theory of Individual Choice Under Uncertainty: Theory, Evidence, and New Directions*. Cambridge University Press, Cambridge, UK, 1992.
- [39] H. Moulin. *The Strategy of Social Choice*. North Holland, New York, 1983.
- [40] J. V. Neumann and O. Morgenstern. *Theory of Game and Economic Behavior*. Princeton University Press, Princeton, NJ, USA, 1947.
- [41] E. Ostrom. A behavioral approach to the rational choice theory of collective action. *American Political Science Review*, 92:1–22, 1998.
- [42] P. K. Pattanaik and Y. Xu. On ranking opportunity sets in terms of freedom of choice. *Recherches Economiques de Louvain*, 56:383–390, 1990.
- [43] B. Peleg. *Game Theoretical Analysis of Voting in Committees*. Cambridge University Press, Cambridge, UK, 1984.
- [44] B. Peleg. Effectivity functions, games and rights. In J. Laslier, M. Fleurbaey, N. Gravel, and A. Trannoy, editors, *Freedom in Economics: New Perspectives in Normative Analysis*, pages 116–132. Routledge, London, 1998.
- [45] E. Picavet. *Choix Rationnel et Vie Publique*. Presses Universitaires de France, Paris, 1996.
- [46] E. Picavet. Rational agreement on human rights in a pluralistic setting. *Rechtsphilosophische Hefte*, 7:109–135, 1997.
- [47] C. Puppe. Freedom of choice and rational decisions. *Social Choice and Welfare*, 12:137–154, 1995.
- [48] C. Puppe. An axiomatic approach for 'preferences for freedom of choice'. *Journal of Economic Theory*, 68:174–199, 1996.
- [49] M. K. Richter. Revealed preference theory. *Econometrica*, 34:635–645, 1966.
- [50] J. E. Roemer. *Theories of Justice*. Harvard University Press, Cambridge MA, 1996.
- [51] P. A. Samuelson. Consumption theory in terms of revealed preferences. *Econometrica*, 15:243–253, 1948.
- [52] J. L. Savage. *The Foundation of Statistics*. Wiley, New York, 1954.
- [53] A. K. Sen. *Collective Choice and Social Welfare*. San Francisco, Holden Day, 1970.
- [54] A. K. Sen. Choice functions and revealed preference. *Review of Economics Studies*, 38:307–317, 1971.

- [55] A. K. Sen. Rational fools: A critique of the behavioural foundations of economic theory. *Philosophy and Public Affairs*, 6:317–344, 1977.
- [56] A. K. Sen. Freedom of choice: Concept and content. *European Economic Review*, 32:269–294, 1988.
- [57] A. K. Sen. Welfare, freedom and social choice. *Recherches Economiques de Louvain*, 56:451–486, 1990.
- [58] A. K. Sen. Welfare, preferences and freedom. *Journal of Econometrics*, 50:15–29, 1991.
- [59] H. Sidgwick. *The Methods of Ethics*. Oxford University Press, Oxford UK, 1907. 7 ème édition.
- [60] R. Sugden. Reciprocity: the supply of public goods through voluntary contributions. *Economic Journal*, 1984.
- [61] B. Walliser. Rationalité instrumentale et rationalité cognitive. Université du Québec à Montréal, Cahiers d'Epistémologie no. 9212, 1992.
- [62] A. Wolfelsperger. Sur l'existence d'une solution 'kantienne' du problème des biens collectifs. *Revue Economique*, 50:879–902, 1999.