UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LES MOTIVATIONS NON RATIONNELLES DANS LA VIE SOCIALE. CONTRIBUTION À UNE THÉORIE DE L'ACTION COLLECTIVE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN PHILOSOPHIE

PAR

LEARRY GAGNÉ

AOÛT 2006

REMERCIEMENTS

Cette thèse repose avant tout sur les généreux conseils et les encouragements de tous les instants de mes deux co-directeurs, Robert Nadeau (UQAM) et Dominique Leydet (UQAM); je les remercie chaleureusement. Je souligne au passage le support financier qu'ils m'ont apporté à titre d'assistant de recherche, cela m'a grandement aidé à me rendre jusqu'au bout. J'aimerais également remercier Paul Dumouchel (Univ. Ritsumeikan, Kyoto), Bernard Walliser (École Nationale des Ponts et Chaussées, Paris), et Jocelyne Couture (UQAM) pour leurs commentaires élaborés sur des versions préliminaires de la thèse. Des parties de la thèse ont été présentées à plusieurs séminaires du Groupe de Recherche en Épistémologie Comparée (UQAM), ainsi qu'au 72ème Congrès de l'ACFAS, UQAM, mai 2004; j'en profite pour remercier les participants. J'ai bénéficié pour la rédaction, d'une Bourse de fin d'études de la Faculté des sciences humaines, UQAM, en février 2005; ce support financier fut grandement apprécié.

Le second chapitre de la thèse a été publié dans la revue *Philosophiques*, vol. 29, no. 2, 2003. Des versions préliminaires des chapitres 3 et 4 ont été publiés dans les Cahiers du GREC, UQAM, no. 2003-18 et 2004-10.

TABLE DES MATIERES

RÉSUMÉ

Dans les sciences sociales, la théorie du choix rationnel est très bien adaptée pour expliquer et prédire les comportements des agents dans des situations de type marché, où ceux-ci cherchent à maximiser leur utilité en choisissant parmi un ensemble d'options qui s'offre à eux, possédant une valeur tirée d'une échelle commune, par exemple un ensemble de biens affichant un prix. La théorie du choix rationnel a la prétention de s'appliquer à une grande variété de phénomènes sociaux; toutefois, nous remarquons que plus la dynamique du phénomène à expliquer s'éloigne de l'idéal du marché, moins cette théorie nous apparaît convaincante. Afin de préserver l'hypothèse de l'agent rationnel, les modèles tirés de la théorie doivent recourir à des hypothèses auxiliaires de comportement qui, bien souvent, correspondent peu à ce qui motive réellement l'agent.

Nous proposons une relecture critique de la théorie du choix rationnel dans de telles situations. Nous prenons comme point de départ la variante de la théorie offerte par Jon Elster, qui comporte deux particularités principales : une distinction nette entre comportements rationnels et non rationnels fondée sur la formation correcte des désirs et des croyances, et l'hypothèse des "motivations mixtes" stipulant que l'agent puisse être motivé par ses émotions, celles-ci entrant en conflit avec les motivations rationnelles. Le sujet principal de notre analyse porte sur les motivations non rationnelles. Nous stipulons que, en plus de la possibilité de comportements irrationnels causés par des désirs ou des croyances mal formées, il existe des situations où l'agent ne veut pas se comporter de façon rationnelle. Nous relions ce genre de comportement aux motivations éthiques, que nous fondons sur les émotions, ainsi qu'aux actions publiques, où le comportement désintéressé peut conférer un gain de réputation et de reconnaissance de la part d'autrui.

Après avoir passé en revue l'approche rationaliste face aux comportements à l'apparence non rationnels (chapitre 1), nous tentons une première application de nos principes dans le champ de la délibération politique. Pour ce faire, nous offrons une critique de la théorie de la démocratie délibérative à la lumière des motivations rationnelles et non rationnelles (chapitre 2). Par la suite, nous abordons la problématique théorique de la rationalité et des émotions, dans le but de fonder notre modèle de l'action collective (chapitre 3). Nous offrons au chapitre 4 une nouvelle manière d'aborder le conformisme aux normes sociales, suivant l'idée de Pierre Bourdieu selon laquelle les agents ne veulent pas paraître se conformer aux normes par pur intérêt. Nous incluons dans notre modèle le "conformisme hypocrite", ainsi que la possibilité d'"équilibre hypocrite" où tous font semblant de suivre la norme de façon désintéressée. Nous terminons notre étude au chapitre 5 en abordant un autre champ social où nous croyons que notre modèle pourrait s'appliquer de façon

avantageuse, soit les relations de pouvoir en organisation. Nous nous attardons sur l'autorité légitime, une forme de pouvoir qui n'est reconnue que si elle ne s'exerce pas ouvertement comme pouvoir. Nous y abordons les conséquences que notre approche pourrait avoir sur l'étude des hiérarchies et des relations humaines en entreprise.

Mots clés : choix rationnel, action collective, normes sociales, émotions

INTRODUCTION

Depuis un demi-siècle, la théorie du choix rationnel connaît beaucoup de succès en sciences sociales. Dans l'explication des comportements des agents en société, la théorie du choix rationnel s'accomplit à son mieux de sa tâche dans des champs sociaux de type marché, où les agents tentent consciemment de maximiser leur intérêt et où les objets convoités ou échangés peuvent être aisément comparables (ils ont un prix, par exemple). Même si l'on sait que, dans un marché quelconque, les individus ne se comportent pas tous de manière rationnelle, on peut supposer, sans perte cruciale d'information, qu'ils le sont effectivement; c'est de cette façon que la science économique aborde les marchés. Par contre, les modèles rationnels s'avèrent moins convaincants lorsqu'ils abordent des champs "hors marché" où les individus sont mus par des émotions, des valeurs et des normes plutôt que directement par leur intérêt bien calculé. Le "paradoxe du vote" en est l'illustration la plus fameuse : selon la théorie du choix rationnel, la participation à une élection populaire devrait être quasi-nulle, puisque la probabilité pour un agent que son vote fasse une différence est très faible, alors que les coûts (temps, déplacement, etc.) sont relativement élevés. Or, nous savons que la plupart des citoyens se présentant aux urnes sont motivés plus par des vertus civiques que par leur intérêt égoïste. Les modèles rationnels de vote, s'ils veulent préserver l'hypothèse de rationalité, doivent postuler des préférences et des fonctions d'utilité particulières qui ont généralement peu à voir avec les motivations avouées des électeurs réels. Sans vouloir nier la pertinence de ces modèles, en autant qu'ils soient bien construits, ils nous apparaissent souvent étriqués, en tout cas moins convaincants que les modélisations de marché.

La théorie du choix rationnel se fonde sur un individualisme méthodologique; l'élément de base de ses modèles est l'*agent*, que l'on considère comme étant doté de désirs (ou de préférences), de croyances et d'intentions. Selon ses croyances, on demande à l'agent d'ordonnancer ses préférences, soit en fournissant une liste ordonnée, soit en attribuant à chacune une valeur subjective d'utilité. Le *choix rationnel* consiste alors pour l'agent à satisfaire sa préférence la plus élevée, en adoptant les moyens les plus efficaces pour y parvenir, ce que nous appelons également la rationalité *instrumentale*. Cette motivation rationnelle constitue à proprement parler l'*hypothèse de rationalité* : l'agent est supposé

chercher à réaliser sa préférence la plus élevée. Il faut immédiatement noter qu'il ne s'agit que d'une hypothèse; il existe d'autres motivations possibles dont nous traiterons plus loin. Toutefois, l'hypothèse de rationalité est essentielle à la théorie du choix rationnel, elle fait partie pour ainsi dire de son noyau dur.

Dans la version axiomatique traditionnelle de la théorie, l'agent est supposé entretenir des croyances parfaites sur le monde, ainsi qu'un ensemble de préférences complet, continu, et transitif. Puisque dans cette thèse, nous nous intéressons avant tout aux motivations rationnelles, nous mettrons en veilleuse les questions concernant les croyances pour nous pencher sur les préférences[1]. Une préférence est une relation entre deux éléments de la situation de choix; une préférence est dite forte si l'agent préfère A à B, et faible si l'agent soit préfère A à B ou est indifférent entre les deux. Dans un ensemble ordonné (ou à ordonner) d'options de choix, on désigne également par "préférence" les éléments individuels. En résumé, l'ensemble des préférences est *complet* si l'agent entretient une préférence ou une indifférence pour chacune des paires possibles dans la situation de choix, il est *continu* lorsque, dans l'ordonnancement (A, B, C), il existe une loterie portant sur les options A et C telle que l'agent s'avère indifférent entre celle-ci et B[2], et il est *transitif* lorsque, pour un ensemble (A, B, C), si A est préféré à B et B est préféré à C, alors A est préféré à C. Ces trois propriétés servent avant tout à permettre l'élaboration de modèles mathématiques de la décision. A partir de telles préférences et de l'hypothèse de rationalité, nous sommes en mesure d'élaborer des modèles rigoureux de décision nous permettant de déterminer ce que l'agent a fait, fera, ou devrait faire, selon que nous nous situons dans une perspective explicative, prédictive, ou normative.

La théorie du choix rationnel est souvent considérée comme une théorie de la décision impliquant un seul agent dans une situation particulière, alors que la théorie des jeux, une

---

[1] Dans la plupart des scénarios, la théorie peut très bien s'accommoder des croyances imparfaites en assignant à l'agent des probabilités subjectives sur la possibilité de certains états du monde, sans pour autant s'éloigner de la rigueur mathématique.

[2] Supposons que l'agent préfère la pomme à l'orange, et l'orange à la banane. On lui présente le choix entre obtenir l'orange avec certitude, et participer à une loterie entre la pomme et la banane. Lorsque les chances de gagner la pomme sont fortes, il préférera prendre sa chance à la loterie. Inversement, si les chances sont faibles, il préférera se contenter de l'orange. Il sera indifférent entre les deux possibilités précisément au point de culbute. Son ensemble ordonné de préférences (P, O, B) est ainsi continu.

branche de la théorie du choix rationnel, porte sur les interactions stratégiques entre plusieurs agents. Mais il y a plus que le nombre d'agents impliqués. Nous pouvons catégoriser le choix rationnel en théories restreintes et étendues (Elster 1983, ch. 1). La théorie restreinte se veut axiomatique et offre un traitement mathématique du choix tout en faisant abstraction de l'"état d'esprit" de l'agent, alors que la version étendue cherche à fouiller cet état d'esprit, soit la constitution des désirs, des croyances, et des intentions, quitte à s'éloigner de la modélisation mathématique. La théorie étendue se veut donc plus "philosophique" dans son approche de la rationalité, et c'est cette variante que nous adopterons tout au long de cette thèse. La théorie des jeux se situe clairement dans la variante restreinte, à un point tel qu'il devient difficile de parler d'"agent", au sens où nous l'entendons en général dans les sciences sociales. Un jeu consiste en une matrice de paiements (la forme normale) ou un arbre de décision (la forme séquentielle) dans lequel les possibilités de choix ainsi que les utilités correspondantes sont fixées *a priori*. Les choix rationnels se déduisent uniquement à partir des données du jeu[3]. L'hypothèse de rationalité prend en théorie des jeux la forme d'un axiome : le joueur est toujours strictement motivé par la maximisation d'utilité, et aucune autre motivation n'est permise. Il n'existe pas de contexte social ou de profils psychologiques plus larges que le jeu, pouvant faire en sorte que l'agent ne se comporte pas de la manière prédite par le jeu. Les valeurs d'utilité inscrites dans la matrice représentent absolument tout ce qui motive l'agent, et il ne peut rien y avoir au dehors[4]. Dans la théorie étendue du choix rationnel, des contraintes de nature socio-psychologiques peuvent influencer le comportement individuel, par exemple lors

---

[3] Le Dilemme du prisonnier est une cause célèbre en théorie des jeux justement parce que le choix rationnel préconisé par le modèle ne correspond pas au meilleur choix des agents tout compte fait (en termes techniques, la solution est en équilibre de Nash mais ne constitue pas un optimum de Pareto). D'innombrables propositions ont été avancées pour permettre aux joueurs d'atteindre la "vraie" solution. La seule vraiment satisfaisante met en scène une répétition de longueur indéterminée du jeu (la stratégie "donnant-donnant"). Mais dans le jeu joué une seule fois, il n'y a vraiment rien à faire. Seules deux avenues sont envisageables pour en arriver à la solution coopérative : soit que l'on introduise des incitatifs afin de rendre la solution coopérative rationnelle, ou soit que l'on établisse des institutions externes permettant aux joueurs de passer contrat et de s'engager à coopérer même s'ils peuvent obtenir mieux en faisant défection. Dans le premier cas, nous transformons le jeu en autre chose que le Dilemme du prisonnier, et dans l'autre, nous sortons du cadre de la théorie des jeux.

[4] Hollis (1996: 72-79) souhaite que la théorie des jeux tienne compte du contexte et des dispositions préalables des joueurs, mais il ne nous dit pas comment nous pourrions y arriver. Contre un tel projet, nous pourrions soulever que, d'abord, les utilités contiennent déjà tout le contexte de choix, et ensuite, que la théorie des jeux est une abstraction mathématique d'interactions, et non pas une théorie sociale.

de changement de préférences, comme nous le verrons à l'instant. Puisque notre sujet principal concerne les motivations, nous n'aurons pas recours à la théorie des jeux comme tel, sauf pour évaluer les tentatives de modélisation de certains comportements non instrumentaux (chapitre 1).

Au cours de cette thèse, nous reviendrons à maintes reprises sur notre conception de la théorie étendue du choix rationnel et de ses limites. Dans les lignes qui suivent, nous proposons une critique de la théorie restreinte tout en tentant de dégager les pistes qui nous mèneront vers une théorie étendue. Prenons comme point de départ les préférences. Comme nous l'avons relevé précédemment, la théorie restreinte exige que l'ensemble ordonné des préférences soit complet, continu et transitif. Cela constitue une exigence très forte à l'endroit des agents. Même si la théorie du choix rationnel, comme bien d'autres théories du comportement, part d'agents idéaux, il est évident que plus l'idéal-type se veut réaliste, plus le pouvoir explicatif et prédictif de la théorie sera grand, au prix toutefois d'une plus grande complexité. Peut-on affaiblir un tant soit peu les axiomes des préférences tout en conservant la simplicité des modèles de la théorie restreinte ? La complétude ne pose pas trop de problèmes à la modélisation. Face à des préférences incomplètes, le modélisateur peut la plupart du temps "remplir les blancs" de l'ensemble de préférences sans changer les paramètres de choix. La continuité se voit violée dans les cas de préférences dites "lexicales" où un critère de choix domine absolument les autres, de telle manière que certains biens n'ont "pas de prix" par rapport à d'autres[5]. Dans de tels cas, une fonction d'utilité ordinaire ne peut être formulée. La solution consiste alors à attribuer un prix extrêmement élevé (mais fini) à ces biens qui n'ont pas de prix.

Dans le cas des axiomes de complétude et de continuité, leur violation ne constitue pas au premier regard des instances d'irrationalité. Il nous semble tout naturel qu'un agent ne puisse pas entretenir en tout temps un ensemble complet et continu de préférences, tout comme des croyances parfaites. Pour cette raison, nous pouvons compléter les préférences de l'agent

[5] Reprenons l'ensemble ordonnée de notre panier de fruits (P, O, B), cette fois-ci avec une banane empoisonnée. Il est possible que, pour un agent normalement constitué, il n'existe aucune probabilité positive, aussi petite soit-elle, de mourir d'empoisonnement face à laquelle il exprimera une indifférence avec l'orange. En d'autres termes, sa vie n'a pas de prix pouvant être exprimé en valeur de pommes ou d'oranges, et ses préférences sont ainsi non continues.

pour fins de modélisation sans trop de soucis. La violation de la transitivité semble par contre nous révéler une véritable irrationalité, au sens d'un défaut de raisonnement. Si l'agent préfère A à B, et B à C, il nous semble absurde qu'il puisse préférer C à A. Et pourtant, il est facile d'imaginer que l'agent puisse changer ses préférences avec le temps et finir par réellement préférer C à A. La théorie restreinte ne peut adéquatement traiter de la temporalité des préférences. Dès que l'on permet à l'agent de changer ses préférences, l'intransitivité devient impossible, car toute découverte d'une intransitivité pourra être excusée par un changement d'attitude face aux options présentes (Rosenberg 1992: 122). Il n'y a ainsi plus de critère de stabilité des préférences, essentielle à la modélisation. De l'autre côté, une hypothèse de préférences stables, souvent admise en économie, empêche l'agent de changer ses préférences, et le modèle sera pris en défaut si l'on observe que l'agent a effectivement changé d'opinion. Comme les préférences sont données (ou révélées) en théorie restreinte, il n'y a pas de raisons derrière les préférences, et il ne peut y en avoir derrière les changements de préférences non plus (Anand 1987: 193). Cette tension entre l'exigence de transitivité, que Elster (1983: 6) considère comme l'axiome fondamental définissant les préférences rationnelles, et le phénomène commun des changements de préférences, constitue un sérieux problème pour la théorie restreinte.

En ordonnant ses préférences, l'agent peut leur attribuer une valeur subjective d'utilité. L'utilité est dite "ordinale" si elle ne représente qu'un index d'ordre, de l'option la plus préférée à la moins préférée, et "cardinale" si la valeur représente une intensité de préférences[6]. Étant donné que les échelles d'utilité sont subjectives, les comparaisons directes entre agents s'avèrent impossibles, à moins que l'on trouve une mesure commune, comme les prix sur le marché, tout en émettant l'hypothèse auxiliaire que l'utilité de l'argent est la même pour tous. L'attribution de valeurs d'utilité s'effectue hors rationalité; la théorie ne s'exprime pas sur ce qui est valide ou non comme utilité (Demeulenaere 1996: 243). Le principal problème théorique soulevé par le concept d'utilité réside dans la commensurabilité. L'utilité ramène toutes les préférences de l'agent sur une même échelle, mais quelle est la signification de cette échelle ? Lorsque les préférences sont suffisamment homogènes, comme des prix sur le marché

[6] Si l'on a recours aux utilités cardinales, les préférences doivent absolument respecter les trois axiomes.

par exemple, l'échelle d'utilité a intuitivement un sens, mais plus les préférences se révèlent hétérogènes, plus l'échelle apparaît abstraite et mystérieuse (Hollis 1996: 45; nous revenons en détail sur la commensurabilité au chapitre 1).

L'hypothèse de rationalité stipule que l'agent maximise son utilité, c'est-à-dire qu'il agit en fonction de sa préférence la plus élevée. Il est redondant d'affirmer que l'agent "vise" la maximisation d'utilité, car la maximisation est implicite dans l'acte même d'ordonnancement des préférences (Hausman 1992: 18). L'agent n'est donc pas rationnel parce qu'il maximise son utilité; au contraire, c'est parce qu'on le considère rationnel (par hypothèse) que l'on s'attend de lui qu'il maximise son utilité (Demeulenaere 1996: 240). 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 exhibant une préférence pour la pauvreté dépend entièrement de ce que l'on veut modéliser.

Afin de contourner les problèmes posés par les préférences subjectives, on peut recourir, en théorie restreinte, à l'approche des "préférences révélées". Cette approche se veut une méthode empirique de détermination des préférences individuelles. Il s'agit d'observer les choix réellement effectués par les agents, et, partant simplement de l'axiome de transitivité et de l'hypothèse forte de rationalité, nous pouvons reconstruire les buts de l'agent ainsi que son ensemble de préférences. De cette manière, les préférences se retrouvent complètes et continues de façon triviale, et l'agent maximise toujours son utilité. Toutefois, cette approche ne règle pas les problèmes soulevés plus haut concernant les choix se révélant intransitifs. Le recours par les chercheurs aux préférences révélées est motivé par un désir de pratiquer une science empirique et objective, et ainsi à ne pas chercher à attribuer des états mentaux aux agents, mais il semble que dans les cas d'intransitivité, si l'on prétend que l'agent est rationnel, l'explication se trouve dans le changement de préférences sans corrélat avec un changement de

situation, donc dans ses états mentaux. Du moment que le seul axiome des préférences révélées est violé, ce qui n'est pas rare, il faut revenir aux préférences subjectives (Rosenberg 1992: 119-21). De plus, pour que l'observateur puisse obtenir un ensemble complet et continu, il faut que le sujet effectue tous les choix nécessaires, sinon il devra inférer le reste, et donc dériver de nouveau vers le subjectif (Demeulenaere 1996: 253, 263). Finalement, les préférences révélées ne nous permettent pas de distinguer entre choix intentionnels et non intentionnels (Hausman 1992: 20-22).

Une approche alternative fut proposée par Stigler et Becker (1977), et reprise par Becker tout au long de sa carrière. Elle consiste à fixer un ensemble général de préférences des agents, et à expliquer tout comportement apparemment causé par un changement de préférence par des variations de prix et de revenus d'investissement. Les auteurs avancent une "nouvelle théorie des choix du consommateur", soit une fonction générale de production où les agents produisent des commodités en investissant leur temps, leur argent, etc., au lieu d'une simple fonction d'utilité passive portant sur un panier de commodités. Si l'agent, qui préférait initialement A à B, inverse sa préférence, ce n'est pas par pur changement de goût, mais parce que le prix de B est devenu plus intéressant, ou encore que l'investissement dans B est devenu plus rentable. Selon eux, par exemple, l'effet du marketing est de faire croire que le produit annoncé représente un investissement plus rentable que ceux de ses compétiteurs (Stigler, Becker 1977: 83-87). La stratégie de supposer une variation de prix plutôt que de préférence renforce l'approche des préférences révélées. En effet, si l'on observe un comportement apparemment intransitif, c'est parce que les prix ont changé durant la période entre les deux décisions contradictoires. Si l'agent ne semble pas maximiser son utilité, c'est que l'observateur n'a pas relevé toute la grille coûts/bénéfices dont l'agent s'est servi[7]. La question du comportement non intentionnel est aussi réglée par l'hypothèse que l'agent n'a pas nécessairement à être conscient qu'il maximise (Becker 1986: 111-112). Becker en vient à expliquer les mariages, les divorces et la procréation comme des décisions rationnelles sur un "marché" d'époux ou de style de vie où tous les "biens" présents offrent un certain rendement en termes d'utilité (Becker 1986).

---

[7] Coleman (1990: 18) reprend la même hypothèse dans l'explication sociologique.

Avec les préférences révélées et stables, la théorie restreinte du choix rationnel prend les allures d'une théorie empirique des comportements sociaux, en évitant le plus possible de formuler des spéculations sur ce qui se passe réellement dans la tête de l'agent. Tout en souhaitant demeurer du côté de l'empirisme, Simon a critiqué cette façon de faire. Il qualifie cette approche de "rationalité substantielle", et l'attribue à la théorie néoclassique en économie. Puisque les caractéristiques psychologiques de l'agent sont entièrement fixées par l'hypothèse forte de rationalité et les préférences stables, on peut donc prédire les choix simplement à partir des données du monde, via la fonction d'utilité. Cela suppose bien sûr des croyances parfaites et une capacité de calcul sans failles. Principalement, sa critique porte sur les nombreuses hypothèses auxiliaires que le modélisateur doit apporter afin de préserver l'hypothèse de rationalité, dans des situations qui s'éloignent de la logique pure de marché. En d'autres termes, il s'agit d'encadrer les paramètres de la situation de telle manière que l'agent puisse être conçu comme pleinement rationnel. Pour Simon, on ne peut appliquer le principe de charité à ces hypothèses auxiliaires, comme on le fait avec l'hypothèse de rationalité; le modélisateur se doit d'expliquer l'origine de ces hypothèses, ou, le cas échéant, abandonner la rationalité (Simon 1987). En proposant la "rationalité procédurale", Simon consent à scruter l'esprit de l'agent afin de pouvoir déterminer ses procédures de décision; on se rendra ainsi compte que l'agent ne maximise pas toujours, que ses croyances sont la plupart du temps incomplètes, et que ses capacités de calcul sont limitées. Le passage de la rationalité substantielle à la rationalité procédurale correspond à un passage du raisonnement déductif axiomatique à une exploration empirique de la pensée (Simon 1982: 442). Simon confère à la psychologie béhavioriste le soin de déterminer les processus de décision des agents. Ce que l'on nomme aujourd'hui l'"économie béhavioriste" porte surtout sur les capacités cognitives et computationnelles des agents[8].

Ces approches alternatives servent principalement à établir des balises plus réalistes à l'attribution de désirs et de croyances lors de la conception de modèles. Cependant, l'hypothèse de la rationalité instrumentale y est constamment maintenue. La rationalité évolutionnaire (Weibull 1998; Danielson 2004) offre une approche radicalement différente. En réponse à la complexification toujours croissante de la théorie du choix rationnel qui s'intéresse de plus en

[8] On retrouve des discussions sur la rationalité, le béhaviorisme et la cognition chez Rabin (1998) et Ostrom (1998) pour les sciences économique et politiques, Schmidt (2001) et Walliser (1989) pour la théorie des jeux, et Livet (2001) pour les sciences sociales en général.

plus aux mécanismes mentaux, la rationalité évolutionnaire élimine tout recours au raisonnement de l'agent en lui attribuant des stratégies toutes faites, et en établissant un mécanisme exogène d'ajustement des stratégies selon les résultats obtenus à chaque itération du modèle. Les agents deviennent ainsi des "hôtes" de stratégies plutôt que des "créateurs", et l'hypothèse de rationalité se retrouve extériorisée dans le mécanisme d'ajustement.

Malgré tous ces remarquables développements, une importante lacune demeure : la théorie du choix rationnel s'intéresse peu aux *motivations* des agents. Elle suppose l'agent toujours rationnel, exception faite de la théorie évolutionnaire, pour qui les motivations n'ont pas d'importance. Comme nous l'avons relevé au début, les motivations rationnelles se marient très bien aux situations de marché; mais avec les valeurs et les normes, ce n'est pas si sûr. Nous sommes convaincus que la théorie du choix rationnel constitue actuellement l'approche la plus intéressante en sciences sociales; elle constitue certainement la théorie la plus avancée permettant d'expliquer les phénomènes collectifs par les actions individuelles sous-jacentes. Nous sommes également d'avis que cette théorie ne peut traiter adéquatement des motivations non instrumentales. On retrouvera dans le premier chapitre une justification détaillée de cette position. En résumé, bien que la théorie du choix rationnel puisse, en toute légitimité, modéliser les motivations non conséquentialistes comme si elles étaient rationnelles, nous soutenons que de telles motivations peuvent avoir des effets spécifiques sur les interactions entre agents, effets qui se perdent lorsqu'on rationalise ces motivations pour fins de modélisation. Par ailleurs, il ne nous apparaît pas possible d'affaiblir l'hypothèse de rationalité, au sens d'admettre des agents systématiquement non rationnels dans un modèle rationnel[9], sans affaiblir du même coup la théorie dans son ensemble, au point de la rendre inutilisable. La question de fond posée dans cette thèse est la suivante: comment prendre au sérieux les motivations non rationnelles tout en conservant le plus possible les postulats de la théorie du choix rationnel ?

La stratégie que nous adoptons afin de répondre à cette interrogation ne consiste pas à chercher à réformer la théorie du choix rationnel, mais plutôt à élaborer une théorie qui

---

[9] La plupart des modèles admettent, implicitement ou explicitement, la possibilité de comportements irrationnels, mais seulement comme de rares accident de parcours.

pourrait référer à la fois aux motivations rationnelles, à l'aide justement de la théorie du choix rationnel, et aux motivations non rationnelles, à l'aide d'un modèle qui reste à construire. C'est la stratégie adoptée notamment par Elster (1999), qui fonde ce second type de motivations sur des réactions de nature émotionnelle, tout en établissant une distinction nette entre émotion et rationalité. Tout au long de cette thèse, nous allons suivre cette voie que Elster a largement contribué à défricher, tout en y apportant des modifications significatives. Nous avons choisi de nous concentrer sur trois problématiques sociales propices à exhiber simultanément les deux types de motivations, soit la formation et l'expression des préférences dans l'exercice démocratique, le conformisme aux normes sociales, et la légitimation de l'autorité en contexte organisationnel. Nous adopterons dans les trois cas une méthodologie similaire, qui consistera d'abord à examiner les théories existantes de l'action collective, ensuite à élaborer un modèle intégrant nos deux types de motivations, et pour terminer, à formuler des hypothèses théoriques sur les conséquences des interactions entre agents rationnels et non rationnels.

La psychologie et la sociologie semblent être les disciplines les plus aptes à pouvoir contribuer aux fondements des motivations non rationnelles, et ainsi à contrebalancer la théorie du choix rationnel, en autant qu'elles relèvent elles aussi de l'individualisme méthodologique (Elster 1989; Kolm 1986). Les prémisses psychologiques que nous avons retenues sont inspirées avant tout de la théorie de Elster. Nous offrons toutefois ce que nous croyons être des améliorations appréciables conduisant à un fondement analytique plus satisfaisant. Nous faisons notamment une très large part au phénomène de duperie de soi (*self-deception*). Notre contribution la plus originale se situe du côté de la sociologie. Nous avons choisi de reprendre, dans un cadre individualiste, certains éléments de la sociologie de Pierre Bourdieu, notamment en ce qui a trait aux profits de reconnaissance au sein du groupe, à l'euphémisation des motivations, et au pouvoir symbolique. A notre connaissance, personne n'a tenté un tel rapprochement de la sociologie de Bourdieu à la rationalité et à l'individualisme méthodologique. Les critiques de la part des partisans de la rationalité envers Bourdieu ont toujours porté sur son système dans son ensemble, qui, effectivement, demeure largement incompatible avec la rationalité (Alexander 1995; Van den Berg 1998). Notre approche consiste plutôt à extraire certains mécanismes sociaux de son système qui, à notre avis, offrent des perspectives novatrices face aux problèmes sociaux qui nous intéressent, quitte à faire

violence à sa théorie générale. L'intérêt principal de notre modèle motivationnel repose sur une idée au fond simple, mais qui aura d'importantes conséquences sur les modèles rationnels existants, ainsi que sur la façon d'aborder bien des problèmes sociaux : nous postulons que certains agents n'ont pas seulement une disposition à se comporter de façon non rationnelle, ils *veulent* se comporter ainsi. Nous postulons également que, par une logique de l'action collective incluant de tels agents, certains agents rationnels *veulent ne pas paraître* rationnels. Comme nous le verrons dans cette thèse, ce modèle fournira des conclusions surprenantes sur le comportement en groupe, qui diffèrent de ce que les modèles rationnels nous ont offert jusqu'à présent, tout en demeurant le plus possible compatible avec l'approche rationnelle.

Un mot pour terminer sur ce que nous entendons par "rationalité". Nous chercherons tout au long de cette thèse à cerner, de façon théorique, ce qui est rationnel et ce qui ne l'est pas. Nous partons d'une conception "étendue" de la rationalité, ajoutant à l'hypothèse de rationalité certains mécanismes psychologiques dans l'explication du comportement de l'agent rationnel. Nous adoptons une théorie étendue de la rationalité similaire à celle avancée par Elster (1983). L'action est fonction des désirs et des croyances de l'agent. En résumé, Elster exige des désirs qu'ils soient autonomes, sans pour autant nier l'importance de la socialisation dans la formation des préférences. Il exige également des croyances qu'elles soient adéquates par rapport à la situation, sans demander la correspondance parfaite[10]. Ces principes ne sont pas de nature axiomatique; ils sont sujets à interprétation à chaque entreprise de modélisation par le chercheur. Nous ferons usage du terme "rationalité" comme strictement équivalent à la rationalité instrumentale. On retrouve un peu partout dans la littérature rationaliste des rationalités alternatives, comme la rationalité expressive, sociale, axiologique, etc. Nous considérons qu'accoler le terme "rationalité" à ces motivations non instrumentales porte à confusion plus souvent qu'autrement. Nous préférons situer ces motivations hors de la rationalité. Pour qu'un comportement soit rationnel, nous posons deux conditions générales : l'agent doit rechercher les meilleurs moyens pour arriver à ses fins, et il doit agir de façon

[10] En fait, la formation des croyances peut être vue comme une action rationnelle : à partir de ses désirs, de ses croyances actuelles et d'une contrainte de temps, l'agent décide s'il veut acquérir plus d'information sur la situation de choix. Toutefois, comme l'agent ne peut connaître à l'avance la valeur de la nouvelle information par rapport au coût temporel, il devra arrêter ses recherches par instinct, pratiquant ainsi une forme de "satisficing" à la Simon.

consciente et intentionnelle. L'action sera non rationnelle si elle viole une de ces deux conditions.

Lors de nos discussions sur la rationalité, nous désirons éviter deux malentendus possibles que nous rencontrons trop fréquemment dans la littérature, surtout lorsqu'il est question de théorie étendue de la rationalité. D'abord, les qualificatifs "rationnel" et "irrationnel" n'ont rien à voir avec un quelconque jugement de valeur. Un agent irrationnel est un agent qui soit ne maximise pas son utilité, soit n'agit pas de façon intentionnelle. Ce n'est pas un agent stupide. Il y a des comportements irrationnels, comme l'amour ou le sacrifice, qui sont parfaitement louables; de la même manière qu'il existe des comportements rationnels répréhensibles. La théorie du choix rationnel n'a pas à porter de jugement en la matière. Nous sommes toutefois conscients de la connotation négative du terme "irrationnel", nous essaierons autant que possible de faire usage à la place du terme "non rationnel", mais ce sont là deux termes équivalents. Le second malentendu concerne la rationalité comme comportement avantageux, sans égard à l'intentionnalité, ce que nous pourrions nommer la "rationalité fonctionnelle". Les théories évolutionnaires de la rationalité font usage d'un tel principe : est rationnel ce qui accroît le *fitness* de l'agent, peu importe s'il le visait consciemment ou pas. Dans sa théorie des groupes sociaux, Hardin (1995) prétend que les membres du groupe maximisent leur intérêt en se conformant aux pratiques du groupe, sans vraiment en être conscients. Nous demeurons réticents à qualifier de "rationnel" un comportement où l'agent ne se sent pas motivé à atteindre de façon efficace le but préconisé par le modèle. Lors de notre discussion sur les émotions, nous rencontrerons des comportements émotifs qui font en sorte que l'agent, de manière impulsive, maximise en quelque sorte son intérêt. Nous persistons à qualifier de telles motivations de "non rationnelles". En bref, nous postulons trois catégories de comportements non rationnels : le comportement *forcé*, soit par la présence d'un seul élément de choix, d'une réaction impulsive (incluant 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. Dans le cadre de cette thèse, il sera avant tout question de la première catégorie.

Sommaire de la thèse

L'objectif général de cette thèse est d'élaborer une critique constructive de la théorie du choix rationnel en contexte d'interaction sociale. L'approche retenue consiste à tenter d'adjoindre une théorie des motivations non conséquentialistes à la théorie du choix rationnel qui elle, se préoccupe de motivations conséquentialistes. Les objectifs secondaires sont, d'abord, une mise à l'épreuve de la démocratie délibérative à l'aide des motivations non conséquentialistes et de la théorie du choix rationnel, et ensuite l'élaboration d'un modèle du conformisme aux normes sociales, ainsi que d'un modèle de l'autorité légitime en organisation, à partir de l'échafaudage théorique proposé. Nous débutons (chapitre 1) par un compte rendu critique des différentes approches rationnelles concernant les motivations non conséquentialistes. Nous appliquons ensuite l'approche retenue à l'étude de comportements rationnels au sein de la démocratie délibérative (chapitre 2). Par la suite, nous entreprenons l'élaboration d'un modèle des normes sociales fondé, notamment, sur le modèle psychologique de Elster, et des éléments de la sociologie de P. Bourdieu (chapitres 3 et 4). Pour terminer, nous reprenons les idées développées tout au long de la thèse, et nous les appliquons à l'étude de la légitimité des relations de pouvoir, particulièrement au sein des organisations (chapitre 5).

Structure de la thèse

Chapitre 1:

La modélisation des comportements non conséquentialistes en théorie du choix rationnel

Le chapitre d'introduction consiste en une discussion épistémologique sur la portée du choix rationnel, qui traversera toute la thèse. Nous tentons de déterminer la place des comportements non conséquentialistes, notamment le respect des valeurs, dans la théorie du choix rationnel. En partant, il faut bien comprendre que, dans la théorie du choix rationnel, il n'y a pas de limites à ce qui peut constituer une préférence ou une valeur d'utilité; pour tout genre de comportement, le théoricien peut faire "comme si" l'agent qu'il modélise maximise son utilité de façon conséquentialiste. Nous exposons deux critiques générales à cette théorie : en *amont*, au niveau du réalisme de l'hypothèse de rationalité attribuée aux agents modélisés, et en

aval, au niveau de la pertinence des résultats du modèle et de sa puissance explicative. Nous soutenons qu'il est possible de répondre aux critiques en amont en soulignant le caractère intentionnellement idéaliste de la modélisation des agents au sein des modèles rationnels. Les critiques en aval nous semblent plus prometteuses. Nous examinons ensuite des modèles rationnels tentant d'intégrer les valeurs. Les modèles standard, intégrant les valeurs directement dans la fonction d'utilité, ou comme préférence, ne posent pas de difficultés *a priori*. Par contre, les modèles non standard cherchent à conserver le caractère non conséquentialiste du respect des valeurs au sein de la rationalité. Selon nous, il en résulte un problème de commensurabilité: soit que les utilités non standard sont commensurables aux utilités ordinaires, mais alors ces premières perdent leur qualités spécifiques, soit qu'elles ne sont pas commensurables, mais alors elles ne peuvent respecter la théorie de l'utilité, et le modèle rationnel ne peut plus tenir.

Nous comparons par la suite deux théories alternatives de la rationalité, qui cherchent à prendre au sérieux les comportements non conséquentialistes : le "modèle rationnel général" de R. Boudon, qui consiste à élargir la notion de rationalité pour y inclure ces comportements, et les "motivations mixtes" de J. Elster, qui consiste, au contraire, à resserrer la notion de rationalité en établissant des limites à ce qui peut être modélisé comme comportement. Dans le reste de la thèse, nous retenons la théorie de Elster. Elle nous semble plus précise, notamment en ce qui a trait à l'irrationalité et au rôle des émotions, donc plus utile en vue de l'élaboration de modèles. Son principal désavantage est, bien entendu, qu'elle est loin d'être aussi claire et élégante que la théorie classique du choix rationnel. La proposition de Boudon nous semble trop vague pour pouvoir adéquatement servir à la modélisation; dans ses propres tentatives, il tend à confondre comportements conséquentialistes et non conséquentialistes. En conclusion, nous sommes d'avis que, pour pouvoir traiter convenablement des comportements non conséquentialistes, il nous faut sortir du cadre rationnel sans pour autant l'abandonner.

Chapitre 2:
La délibération circonstancielle en théorie démocratique

Tout comme la théorie du choix rationnel, la démocratie délibérative a recours à des agents idéalisés afin de pouvoir modéliser des situations d'interactions sociales. Mais là s'arrête la comparaison, car les citoyens en démocratie délibérative sont supposés impartiaux, raisonnables et tolérants. Une autre différence est que la démocratie délibérative se veut moins axiomatisée que la théorie du choix rationnel. La question que nous nous posons dans ce chapitre est la suivante : n'y aurait-il pas lieu d'intégrer des agents rationnels dans les modèles de démocratie délibérative ? Nous considérons, dans ce que nous nommons la "délibération circonstancielle", des agents rationnels au sens elstérien, pouvant être mus par des motivations non conséquentialistes, qui, selon les circonstances, pourraient se comporter comme les citoyens des modèles délibératifs. Il est de notre avis que des notions telles que le "consensus" et la "force du meilleur argument" ne sont pas suffisamment appuyées par des mécanismes sociaux; nous souhaitons que la délibération circonstancielle puisse contribuer à y remédier.

La première partie consiste en une analyse des différentes théories délibératives, à la lumière de deux phénomènes individuels chers à la théorie du choix rationnel, la formation et l'expression des préférences. Dans le premier cas, nous relevons que les théories délibératives ne donnent pas suffisamment d'importance à la pression du groupe, la persuasion et l'autonomie des préférences. Aussi, la modélisation des citoyens comme préférant la résolution des conflits nous apparaît trop *ad hoc*, au même titre que l'instrumentalisation des valeurs en choix rationnel. Dans le second cas, nous nous intéressons au caractère substantiel des procédures délibératives, c'est-à-dire l'influence des présupposés moraux de la théorie elle-même sur l'expression des préférences, notamment concernant les différents principes d'égalité. Dans la seconde partie, nous offrons des alternatives tirées de la rationalité : comment des agents mus par des valeurs, et sujets à la pression populaire, peuvent être amenés à se comporter de façon raisonnable, les avantages du marchandage sur la délibération dans certaines situations, le recours à des préférences stratégiques dans le but de contrer des propositions moralement inacceptables, et les effets pervers des procédures de décision démocratiques. En introduisant la possibilité de comportements non conséquentialistes chez l'agent rationnel, dans ce cas-ci par les valeurs du citoyen raisonnable, nous croyons être en mesure de rapprocher ces deux théories que sont la démocratie délibérative et le choix rationnel.

Chapitre 3:
Les fondements rationnel et émotif des normes sociales

L'objectif de ce chapitre est de fournir une fondation motivationnelle à notre modèle des normes sociales qui sera développé dans le chapitre suivant. Nous procédons principalement à une refonte du modèle des "motivations mixtes" de Elster. Pour Elster, le respect des normes sociales repose sur l'anticipation de honte, causée par le mépris d'autrui, en cas de manquement à la norme. Nous postulons, à l'aide de diverses théories psychologiques, une continuité entre les émotions de honte et de culpabilité, différenciées par leur intensité, et qui peuvent affecter le conformisme aux normes chacune à leur façon. Nous reprenons l'idée fondamentale de Elster que les motivations de nature émotionnelle peuvent non seulement affecter les valeurs d'utilité de l'agent, elles peuvent aussi entraver, voire bloquer, ses capacités de raisonnement rationnel. Nous postulons que la honte, étant de nature plus intense, bloque le raisonnement rationnel et rend l'action non conséquentialiste, alors que la culpabilité, moins intense, permet un raisonnement rationnel, quoique biaisé. Nous postulons également que le non respect de nos propres valeurs est ce qui génère de la honte ou de la culpabilité, selon l'importance de celles-ci pour nous, et que les normes sociales représentent également certaines valeurs partagées par la communauté. Donc, si la norme exprime une valeur perçue par l'agent comme capitale, il anticipera de la honte face à une violation de la norme, et sera porté à s'y conformer de manière non conséquentialiste. Si la valeur est pour lui moins importante, il anticipera de la culpabilité, et pourra s'y conformer rationnellement tout en maintenant publiquement des intentions vertueuses.

Nous reprenons également le modèle des transformations des motivations de Elster, afin d'expliquer comment les émotions peuvent amener l'agent à changer ses motivations intéressées en motivations vertueuses ou raisonnables. Celui-ci peut être conduit à se mentir à lui-même (*self-deception*), à mentir aux autres par anticipation de culpabilité, ou encore, à mentir de façon tout à fait rationnelle, afin de maximiser son utilité. Lorsque l'agent redécrit ses motivations, il peut recourir soit aux excuses ou aux justifications, avec des contraintes sur ce qui peut être exprimé dans les deux cas. En conclusion, nous retenons trois types de

motivations au conformisme aux normes sociales : le conformisme sincère, motivé par les valeurs, le conformisme par souci d'estime, et le conformisme rationnel. Notre modèle, mettant l'accent sur les motivations, ouvre la voie à l'étude politique des comportements, où les processus de décision deviennent tout aussi importants que les résultats.

Chapitre 4:
Le conformisme non rationnel aux normes sociales, sincère et hypocrite

Après avoir établi les bases motivationnelles des normes sociales, nous passons maintenant à l'élaboration de notre modèle des normes sociales. Notre définition est la suivante : les normes sociales sont des règles externes informelles, partagées par les membres de la communauté, et soutenues par un système de sanctions. Elles se fondent sur certaines valeurs que les membres veulent promouvoir pour leur communauté, c'est ce qui distingue les normes sociales des conventions. Lorsqu'un membre considère ces valeurs comme importantes, on dira qu'il a intériorisé la norme; celui-ci a donc une motivation non conséquentialiste à s'y conformer. D'autres, moins attachés à ces valeurs, seront portés à s'y conformer de manière rationnelle. Ce modèle se fonde sur la rationalité elstérienne telle que révisée plus haut, et également sur la théorie du capital symbolique de P. Bourdieu. Nous retenons de Bourdieu le principe qu'une participation désintéressée et immédiate au bien-être collectif sera toujours plus récompensée qu'une action intéressée; donc, pour notre modèle, un conformisme sincère aux valeurs attirera plus d'approbation qu'un conformisme rationnel. Nous postulons, dans la définition des normes sociales, qu'elles contiennent toutes ce principe, que la "bonne manière" de s'y conformer est toujours non rationnelle. Ce qui distingue principalement notre modèle des modèles de choix rationnel, c'est que la *manière* de se conformer aux normes est capitale.

Bien que toutes sortes d'analyses soient possibles, nous avons choisi de nous concentrer sur le conformisme hypocrite, soit l'agent rationnel qui imite les comportements des agents motivés par les valeurs afin d'obtenir l'approbation de ses pairs. Nous examinons comment l'hypocrisie est possible, autant du côté du conformisme que de l'attribution de sanctions. L'étape suivante consiste à élaborer un modèle d'interaction sociale incluant l'hypocrisie. Nous nous penchons d'abord sur un modèle similaire en choix rationnel,

l'"économie de l'estime", pour ensuite offrir notre propre alternative, l'"équilibre hypocrite". Nous faisons appel à une proposition de Bourdieu : étant donné que l'approbation sociale est désirable pour tous, il est possible qu'un groupe entièrement composé d'agents rationnels fassent semblant de respecter les valeurs de la norme, en autant qu'il y ait une "méconnaissance commune" volontaire, où les agents ne savent pas, *et ne veulent pas savoir*, que chacun agit en fait hypocritement. Lorsque la connaissance du conformisme instrumental des agents devient inévitable, l'approbation sociale devient impossible et la structure d'approbation (réputation, statut, etc.) tombe, ainsi que la norme elle-même. Nous illustrons le concept d'équilibre hypocrite par une étude de la pratique du duel.

Sans vouloir rejeter les approches rationnelles des normes sociales, nous considérons que notre modèle comporte des avantages non négligeables. D'abord, la rationalité peut très bien fournir des explications au niveau macro, mais si ce qui intéresse le chercheur, ce sont les *pratiques* guidées par les normes, alors on doit recourir à des postulats de l'agent hors-rationalité car, selon nous, le conformisme *ouvertement* rationnel à une norme ne peut constituer un équilibre viable. Notre modèle nous permet aussi d'éviter le plus possible ce que nous appelons le "fonctionnalisme rationnel", c'est-à-dire l'attribution de rationalité "comme si" à des agents manifestement non rationnels, sous prétexte que le conformisme sincère génère des bénéfices collectifs. Notre modèle permet aussi d'expliquer comment des normes qui ne contribuent pas au bien-être collectif (des normes de vendetta, par ex.) peuvent se maintenir. Le principal défaut du modèle est, bien entendu, qu'il s'éloigne de l'élégance du choix rationnel et devient beaucoup plus difficile à gérer.

Chapitre 5:
La légitimation de l'autorité dans les théories rationnelles du pouvoir

Ce chapitre reprend le modèle motivationnel utilisé dans les chapitres antérieurs et l'applique au champ du pouvoir. En réponse aux typologies complexes du pouvoir que l'on retrouve dans les modèles rationnels, nous proposons une dichotomie générale entre pouvoir brut et autorité légitime. Tout comme dans notre modèle des normes sociales, le conformisme ouvertement rationnel ne constitue pas la meilleure manière de suivre la norme, dans la plupart

des relations de pouvoir, un exercice brut de pouvoir n'est pas acceptable; le dominant doit alors légitimer son autorité. Les ordres d'un dominant passeront beaucoup plus facilement s'il peut convaincre ses sujets que celles-ci émanent de lois, de règles ou de normes sociales établies plutôt que de sa volonté personnelle. Contrairement aux modèles rationnels, qui voient dans la légitimité un critère d'efficacité, nous affirmons que, dans la plupart des cas, la légitimité est une *condition* de l'exercice du pouvoir, à cause des normes sociales en vigueur.

Nous distinguons deux sources de légitimité : la structure informelle, composée des valeurs communes et des normes sociales, et la structure formelle, composée des lois et règlements. Notre conception de la légitimation informelle se fonde sur la théorie du pouvoir symbolique de Bourdieu, et suit d'assez près nos propos des chapitres précédents. Pour Bourdieu, les agents n'ont qu'une connaissance "pratique", intuitive, de leurs valeurs, à tel point qu'ils peuvent potentiellement s'identifier à plusieurs descriptions différentes. La norme sociale, qui se veut une codification plus formelle de ces valeurs, devient en quelque sorte un point focal autour duquel les agents partageant ces valeurs peuvent se rencontrer. Le pouvoir symbolique est le pouvoir de "faire", en les exprimant d'une certaine manière, les valeurs et les normes; il n'est accordé qu'aux agents de réputation élevée. Le recours aux valeurs fait appel aux motivations de nature émotionnelle, et à un certain conformisme non conséquentialiste. Pour ce qui est de la légitimation formelle, nous avons choisi d'étudier la façon dont les théories du management et des organisations abordent le problème du pouvoir, étant donné que ce type de champ social est habituellement régi par des règles strictes. Comme ces règles se situent conceptuellement plus près des conventions que des normes sociales, la motivation derrière l'exercice de l'autorité, et de son acceptation, aura tendance à prendre une forme plus rationnelle. Les dominés sont prêts à rationnellement accepter une relation de pouvoir directement instrumentale, pourvu qu'elle soit invoquée au nom de la firme, et non à celui du dirigeant. L'attachement émotionnel à des valeurs n'est pas ici nécessaire. En conclusion, nous soumettons l'hypothèse que, lorsque des règles formelles sont présentes, la légitimation formelle domine, alors que dans les champs où ces règles sont moins présentes, les valeurs et les normes prennent le dessus comme source de légitimation. Par exemple, dans une firme fortement hiérarchisée, l'exercice du pouvoir a tendance à être rationnel, l'autorité étant justifiée par les règles, alors que dans une firme à la hiérarchie plus souple, adoptant une

philosophie participative pour les employés, la recherche de réputation à travers la représentation des valeurs du groupe, donc l'exercice d'un pouvoir symbolique, prend de l'importance et façonne les relations entre les membres d'une toute autre manière.

CHAPITRE I

LA MODÉLISATION DES COMPORTEMENTS NON CONSÉQUENTIALISTES EN THÉORIE DU CHOIX RATIONNEL

Le principal problème épistémologique soulevé dans cette thèse concerne le statut des comportements non conséquentialistes au sein de la théorie du choix rationnel. La théorie 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 le 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[11], 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. 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. Toutefois, 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. 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, d'où la nécessité de recourir à l'adéquation des moyens aux fins dans la définition de la rationalité (Demeulenaere 1996: 240). 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 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.

[11] 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.

Nous excluons d'emblée des comportements non conséquentialistes les comportements considérés comme irrationnels dans la conception classique de la rationalité, où l'agent ne réfléchit pas, comme les actions impulsives ou les décisions prises complètement au hasard[12]. Nous pourrions qualifier ces genres de comportements de *techniquement* non conséquentialistes, car ils ne visent même pas de buts. Nous portons 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. Les valeurs nous semblent 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 coûts et aux bénéfices potentiels autres que le respect de la valeur. Le second témoin de la petite histoire 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 premier chapitre, nous effectuerons un survol des types de modélisation rationnelle des comportements non conséquentialistes. Nous débutons 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, dans une volonté de prendre au sérieux la distinction entre conséquentialisme et non conséquentialisme, soit les alternatives proposées par Raymond Boudon et Jon Elster.

1.1 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

[12] Bien entendu, le recours au hasard peut s'avérer rationnel lorsque l'agent, après mûre réflexion, attribue la même valeur maximale d'utilité à certaines alternatives. Ici, la rationalité peut conduire l'agent à éliminer les alternatives indésirables et ainsi à restreindre les possibilités de choix. Voir Elster (1989) sur le recours rationnel à des mécanismes aléatoires de décision.

instance, la question des comportements non conséquentialistes ne s'y pose pas, car on peut toujours faire "comme si" ils étaient conséquentialistes, en autant qu'ils ne soient pas manifestement irrationnels. Dès lors, la question de la portée empirique de la théorie s'impose. 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, au double sens qu'il est réellement motivé par une norme de rationalité, et qu'il possède la compétence intellectuelle requise pour calculer ses fonctions d'utilités. 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. 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. Étant donné qu'en choix rationnel, la rationalité de l'agent est idéalisée, elle s'en trouve peu affectée par la première critique. 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). La théorie du choix rationnel a tendance à prendre plus au sérieux la critique en aval. On peut classer ses variantes en deux catégories : les théories générales et spéciales (Goldthorpe 2000: 123-5). Avec les théories générales, on peut se permettre de modéliser à peu près n'importe quelle situation sociale; on pense à l'"impérialisme économique" de l'École de Chicago, représentée par l'oeuvre de Becker[13] 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 notre survol des modèles d'utilités multiples. En résumé, pour ce qui est de la modélisation des comportements non conséquentialistes, la théorie du choix rationnel, du moins dans ses variantes spéciales, prendra en considération le réalisme des résultats du modèle, et non le réalisme du comportement attribué à l'agent.

[13] Notamment, pour Becker (1986: 111-12), les comportements non optimaux s'expliquent pas des coûts "monétaires ou psychiques" qui ont pu échapper à l'observateur. De plus, l'agent n'est pas nécessairement conscient de son propre effort de maximisation de son utilité.

La théorie peut ramener tout comportement non conséquentialiste en comportement rationnel maximiseur en adoptant des motivations et des préférences *ad hoc*. A 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. 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 se révèle efficace. Par contre, si la situation de l'agent change suffisamment pour que la stratégie traditionnelle cesse d'être 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 une préférence pour la tradition, et alors, "[t]he degree of charity required becomes self-defeating" (Goldthorpe 2000: 129). Expliquer un comportement traditionnel par une préférence pour la tradition apparaît pour Goldthorpe 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. A l'inverse, pour Etzioni (1988: 43-45), la fusion fins-moyens que l'on retrouve dans les décisions morales constitue une raison suffisante de les séparer 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 y renoncera, ou, à tout le moins, qu'il 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[14]. 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

[14] Pettit (1995) nous fournit l'image de la rationalité comme "balises virtuelles" venant limiter nos actions morales lorsqu'elles s'éloignent trop de notre intérêt personnel.

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)[15].

Une valeur peut être modélisée comme une préférence ou comme une contrainte. L'élève veut réussir son examen, mais sans tricher; comment modéliser son honnêteté ? Ben-Ner et Putterman (1998: 20-22) désignent les valeurs comme des préférences sur les manières d'agir (*process-regarding preferences*). Il est possible, selon eux, de les intégrer à un modèle rationnel de deux manières, en les posant soit comme buts de l'action (l'élève désire atteindre le double but de réussir son examen et d'être honnête), soit comme contraintes venant rabaisser, possiblement jusqu'à zéro, l'utilité des alternatives impliquant un manque d'honnêteté. En édifiant l'effet contraignant d'une valeur à un niveau semblable aux contraintes matérielles ou budgétaires, la théorie semble confondre deux catégories distinctes. Une contrainte est par définition indésirable. Nous pouvons souhaiter avoir plus d'argent afin d'élargir notre éventail de choix, mais nous ne souhaitons pas être moins honnêtes afin de nous ouvrir à la possibilité de tricher. L'intériorisation des valeurs signifie qu'elles deviennent pour nous d'authentiques préférences, et non seulement des contraintes (Wolfelsperger 2001: 72-74; Etzioni 1988: 46). La différence devient la suivante : aux prises avec une contrainte budgétaire, je préfère une belle voiture, mais je ne peux me l'offrir, alors que dans le cas axiologique de l'examen, à cause de mon honnêteté, je préfère *ne pas* tricher. On revient donc à l'action guidée par un amalgame de préférences. Il demeure toutefois parfaitement possible, à l'intérieur d'un modèle de choix rationnel, de faire "comme si" les valeurs étaient représentées par des désutilités, puisqu'il n'y a *a priori* aucune limite à ce qui constitue une préférence, ou à ce qui entre dans une fonction d'utilité.

Certains théoriciens 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

[15] 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 fait plus de sens si on lit "...it is what drives up the cost *threshold*". L'impératif rend l'agent plus tolérant des coûts qu'il endosse en agissant de façon morale.

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. Les choix de mots sont importants ("it would seem best..."), car par là, 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 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 vide la théorie de son contenu empirique car elle ne s'intéresse pas aux motivations. Cette "critique épistémologique" de Isaac s'avère "cruciale à tout bon travail empirique". Dans la même veine, Hausman et McPherson affirment : "Even though it may be possible formally to model the commited 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.

Il semble bien que si l'on veuille demeurer fidèle à la théorie du choix rationnel, il y aurait 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 exhiber 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[16] (Ferejohn 2002: 227-28).

1.2 Modèles rationnels des valeurs

Nous allons maintenant nous intéresser plus en détail à des modèles rationnels intégrant les valeurs et les normes, pour voir s'il n'y aurait pas moyen de sauvegarder l'autonomie motivationnelle des valeurs tout en demeurant fidèle à la théorie du choix rationnel. Nous regroupons ces modèles selon qu'ils traitent des valeurs comme des utilités standard, c'est-à-dire conceptuellement conformes en tout points aux utilités matérielles, ou comme des utilités non standard possédant des caractéristiques particulières les distinguant des utilités matérielles.

Modèles d'utilités standard

Deux grands courants de la théorie du choix rationnel proposent une conception entièrement conséquentialiste des valeurs, l'échange social et le néo-institutionnalisme. La théorie de l'échange social est une théorie sociologique fondée sur l'individualisme méthodologique, qui appréhende les relations sociales comme une série d'échanges entre agents, où n'importe quel type de bien, matériel comme symbolique, peut servir de monnaie. Pour George Homans, un de ses fondateurs, le comportement individuel en société se comprend mieux comme le résultat d'un apprentissage béhavioriste : le comportement d'un agent se voit conditionné par les comportements d'autrui. Ce sont les "valeurs" qui provoquent la réaction d'autrui, tandis qu'une "norme" est simplement "a verbal description of behavior that many members find it valuable for the actual behavior of themselves and others to conform" (Homans 1958: 598-600). Bien que la théorie de l'échange social se soit raffinée au fil des ans, la notion d'*attentes réciproques* demeure au coeur de sa conception des normes sociales. Ces normes agissent comme des règles de coordination permettant de maintenir un équilibre social bénéfique. En supposant que l'agent profite personnellement de l'état d'équilibre

[16] 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.

une fois atteint, il se conformera à la norme s'il s'attend que la plupart de ses concitoyens fasse de même. Pour Bicchieri (1993: 230-33), l'adhésion à une norme s'explique par une préférence pour le conformisme, ainsi que la croyance que la plupart des autres s'y conformeront.

Le néo-institutionnalisme est une variante de la théorie du choix rationnel qui se préoccupe des effets des institutions externes, dont les règles formelles et les normes sociales, sur le comportement des agents. Le rôle des institutions est double : elles façonnent la structure au sein de laquelle les agents effectuent leurs choix, par des contraintes et des incitatifs, et elles peuvent influencer directement les préférences des agents (Goodin 1996: 19-20; Ostrom 1986: 6-7). Un équilibre social néo-institutionnaliste peut être vu comme "induit par la structure" (*structure-induced equilibrium*, Shepsle 1989: 137), soit une situation où aucune alternative n'est préférée par tous les autres, à structure institutionnelle donnée. L'échange social et le néo-institutionnalisme sont de proches cousins. Les deux théories abordent les mêmes objets sociaux à l'aide de concepts similaires. Tandis que la première considère les normes comme le produit des relations interpersonnelles, la seconde met plus l'accent sur les effets des normes sur ces relations (Nee, Ingram 1998: 40).

Dans le paradigme échange social/néo-institutionnaliste, une norme sociale n'est rien d'autre qu'une institution servant à coordonner les attentes des membres du groupe, que l'on présume rationnels. Sa fonction est de produire un surplus collectif en incitant les contributions individuelles nécessaires. Le conformisme à la norme est rationnel aussi longtemps que les bénéfices de l'opération pour l'agent dépasse ses coûts de participation, en tenant compte la menace de sanctions planant sur lui en cas de rechignage. De telles normes peuvent conduire à la collaboration dans un Dilemme du prisonnier, par exemple, en réduisant les valeurs d'utilité de la défection au moyen de sanctions. En guise de réponse au problème motivationnel de l'administration des sanctions[17], on suppose que les agents recherchent et valorisent l'approbation des autres, qui se mérite, entre autres, en réprimant les déviants et en récompensant les bons comportements[18]. Nee et Ingram ont recours à une définition de l'intérêt

[17] Les sanctions représentent en soi un problème d'action collective lorsque le geste de sanctionner correspond à un coût net pour l'agent; chacun préfère alors que les autres s'en chargent.

[18] Un système de sanctions n'a pas nécessairement besoin de reposer sur l'approbation d'autrui, même dans le paradigme qui nous intéresse. Un bon exemple est le modèle de Heckathorn (1988,

individuel qui inclut des "biens sociaux" comme la recherche de statut et l'évitement de l'ostracisme; les agents peuvent ainsi toucher des bénéfices "de second ordre" en sanctionnant les déviants (Nee, Ingram 1998: 30-31). Frank (1998: 283) propose dans la même veine les "récompenses internes et non matérielles" comme motivation additionnelle.

Nous ne discuterons pas des modèles issus de la théorie évolutionnaire de la rationalité, pour la simple raison qu'elle ne s'intéresse pas aux motivations des agents, alors qu'il s'agit de notre thème central. La théorie évolutionnaire cherche avant tout à simplifier au maximum le concept de l'agent. En théorie des jeux classique, là où l'on retrouve la conception de l'agent rationnel la plus formelle, être un agent maximiseur signifie, entre autres, détenir une information complète sur le jeu, ainsi que les capacités intellectuelles nécessaires pour pouvoir calculer les fonctions d'utilité et les équilibres. Les nombreux raffinements de ces postulats, au départ irréalistes, apportés au cours des années ont grandement complexifié la théorie[19]. La théorie évolutionnaire nous permet d'explorer une nouvelle voie. Ici, les joueurs 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. Ces joueurs sont tirés au hasard dans une population et placés dans un jeu dans lequel 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[20]. Le principe général de sélection est ce que Bowles et Gintis nomment "réplication 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*

1990). Il conçoit une institution répressive formelle qui va punir tout le groupe même si un seul déviant se fait prendre. Il sera dans l'intérêt des membres du groupe de se surveiller eux-mêmes lorsque les coûts d'administration des sanctions informelles seront surpassés par la désutilité espérée d'être puni si le voisin ne se conforme pas à la norme.

Il ne faut pas oublier non plus que dans un paradigme de choix rationnel, on peut faire "comme si" l'agent exhibait une préférence à sanctionner autrui, même si les coûts directs nous apparaissent supérieurs aux bénéfices (Wolfelsperger 2001: 87).

[19] Tan et Werlang (1988); Binmore (1987); Harsanyi (1990).

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

par un algorithme exogène à l'agent; elles ne sont pas *choisies* par un processus endogène. Il n'y a donc pas lieu de se poser la question si le choix se révèle conséquentialiste ou non.

Modèles d'utilités non standard

Certains théoriciens s'intéressent de plus près aux motivations des agents. Insatisfaits de la modélisation rationnelle *ad hoc* des valeurs, ils assignent des attributs particuliers aux préférences dites "morales" par rapport aux préférences plus matérielles, tout en cherchant autant que possible à 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 les autres sur le marché parce qu'elle représente le meilleur compromis entre ses préférences pour la performance, la consommation d'essence et l'apparence, par exemple. On qualifie ces préférences d'*homogènes* car, partageant la même nature (comme des caractéristiques de voitures, par exemple), on peut les classer sur une même échelle globale d'utilité. L'agent peut sacrifier un peu d'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 des *préférences hiérarchiques*, de leur part, conçoivent les valeurs comme des préférences de second ordre, gouvernant celles de premier ordre. Ces approches génèrent 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 $U = U_A + U_B$, 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, ainsi qu'avec les modèles d'échange social et néo-institutionnalistes[21]. Ce qui nous intéresse dans l'idée d'utilités multiples, 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 équivalons 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 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: 547-48). Ce que Isaac nomme "utilité étendue" correspond à nos utilités additives.

Nous retrouvons une variante du modèle des utilités multiples chez Heath (2001). Tandis que les résultats des actions reçoivent une valeur d'utilité standard, 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 commensurables et

[21] Höllander (1990) nous en fournit un bon exemple. Sa fonction d'utilité comprend deux éléments principaux : la consommation et l'approbation du groupe. Ici, toutefois, 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. 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.

comparables d'une manière qui demeure inexpliquée. 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[22]. 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[23]. 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. En tous cas, les valeurs normatives ne jouissent pas ici de caractéristiques spéciales par rapport aux utilités standard; dans ce modèle, A(a) peut théoriquement exprimer n'importe quoi, des coûts de transaction, des préférences altruistes, l'évaluation de certaines externalités, etc. (Wolfelsperger 2001: 82).

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

---

[22] 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 nécessairement indifférent), sauf si elles sont directement intégrées aux résultats des choix; mais on ne peut les intégrer car alors 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, en 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.

[23] Heath prétend également qu'en présence d'équilibres multiples, la convenance de telle ou telle action peut servir de "point focal" et ainsi rompre l'indétermination. Ici aussi, inclure la valeur de convenance directement dans le jeu ne ferait aucune différence.

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), et voulant dépasser la "dichotomie grossière" de la moralité comme préférence ou comme contrainte (Kuran 1998: 236 n.5), il apparaît nettement, de par sa fonction d'utilité, que la moralité constitue en fait un facteur d'utilité menant, en bout de ligne, 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 retrouvent 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 effectué, son utilité intrinsèque s'estompe, 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.

Sans prétendre à l'autorité finale sur le sujet, il semble que vouloir à la fois des utilité spéciales et une fonction standard d'utilité soit voué à l'échec. Les fonctions d'utilité exigent des paramètres strictement commensurables, sinon elles ne font plus sens. Margolis a reconnu le problème, et il nous suggère de recourir plutôt à une fonction de production. Il distingue les utilités "privées" et "sociales", ces dernières représentant la valeur d'une ressource pour autrui telle qu'évaluée par l'agent (Margolis 1990: 245-46). La fonction de production génère alors des ressources privées ou sociales, dépendamment de leur utilité marginale et de la quantité totale de chacune produite à date. Même si la fonction de production évite d'additionner les deux utilités, elle doit quand même être en mesure de les comparer, et le problème de la commensurabilité refait surface. L'échelle commune d'évaluation suggérée correspond à "the individual's own sense of social values" (Margolis 1991: 355). Pour en comprendre le sens, examinons sa solution au paradoxe du vote. Elle implique la comparaison de l'utilité privée, souvent minuscule, du vote à son utilité sociale telle qu'évaluée par l'électeur : "Hence if Ellie were to attach a numerical value to the advantage to the country of electing what she judges to be the better president, that number could easily be in the [$]billions" (Margolis 1990: 245).

D'où provient ce montant ? Margolis soutient que ce n'est pas en demandant, hypothétiquement, à Ellie quelle somme d'argent serait suffisant pour qu'elle accepte de changer son vote, comme le propose la théorie Von Neumann-Morgenstern de l'utilité (Margolis 1990: 245-46). Il semble plutôt provenir à la fois des normes sociales présentes ainsi que des normes "that rational individuals could be expected to jointly adopt if they could mutually bind themselves" (Margolis 2003: 4). Peu importe le type d'utilité en cause ici, nous savons qu'il ne correspond pas à la théorie VNM de l'utilité[24], alors que l'utilité privée s'y conforme; par conséquence, on ne voit pas très bien comment ces utilités pourraient être adéquatement commensurables. Nous devons en conclure que, dans le traitement des utilités non standard, les fonctions de production souffrent des mêmes faiblesses épistémologiques que les fonctions d'utilités.

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 sur les 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, italiques ajoutées). La satisfaction ou la frustration d'une valeur n'affecte pas l'utilité de l'agent; donc, considérant 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 maximisateur (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 d'abord ses

[24] L'utilité sociale ne correspond pas ici à l'utilité pour l'agent des conséquences sociales de telle ou telle possibilité (ou, plus simplement, sa préférence personnelle pour une meilleure société), mais plutôt à la croyance de l'agent d'une valeur d'utilité existant, d'une certaine manière, objectivement en-dehors d'elle.

préférences de premier ordre selon certaines échelles d'utilité[25], et ensuite procède à l'ordonnancement de ces ordonnancements en vue de la maximisation de valeurs, en supposant que l'agent aspire effectivement à agir moralement (Sen 1982: 100-101). Une telle modélisation des valeurs comme contraintes fortes ne peut fonctionner que si le méta-ordonnancement est *lexicographique*, c'est-à-dire, qu'on a recours aux ordonnancements de premier ordre que pour trancher les situations d'indifférence laissées par le méta-ordonnancement. 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'ordonnancement 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'une valeur d'utilité aux valeurs, dans une fonction d'utilité globale, 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-ordonnancement au moins un ordonnancement lexicographique (c'est pour lui la condition d'existence d'un ordonnancement moral), suivi par des ordonnancements moraux faibles pouvant faire l'objet de compromis, et enfin, des ordonnancements selon l'intérêt personnel. Parmi l'ensemble résiduel de choix laissé par les ordonnancements lexicographiques, l'agent peut recourir à une fonction d'utilité où il assignera des valeurs utilités à ses valeurs, avec un poids correspondant à leur position dans l'ordonnancement global (Ansperger 1998: 205-207). Si ce modèle s'avère correct, c'est-à-dire que le modèle des préférences hiérarchiques ne peut par lui-même gérer le compromis moral, et donc qu'une fonction quelconque d'utilités additives est requise (Ansperger 1998: 206), alors, encore une fois, le modèle des préférences hiérarchiques n'a rien de spécial à offrir[26]. En effet,

---

[25] 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.

[26] Voir la critique similaire de Wolfelsperger (2001: 76-77).

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).

Conclusion

Nous constatons que les modèles rationnels incorporant valeurs et normes 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 utilité morales, ou soit qu'elles s'avèrent incommensurables, et il apparaît alors impossible de pouvoir construire une fonction d'utilité appropriée sans violer 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 puisse 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.

1.3 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, nous allons nous pencher sur deux théories alternatives de la rationalité, qui cherchent à éviter de ramener ces comportements à un moyen pour leur propre fin, et leur donner une place bien à eux. Les deux auteurs que nous proposons adoptent chacun une stratégie à 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 alternative 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 des raisons fortes[27] 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

[27] 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 donc fournissent une base plus solide pour l'explication.

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" en autant que l'agent les croient vraies sur la base de raisons fortes. Contre la rationalité cognitive, Boudon mentionne une "idée reçue", les associations entre 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 soient 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 des raisons fortes de croire que X est juste ou bon. Les raisons axiologiques se fondent sur le "programme de la morale", qui a pour but de promouvoir la "dignité de l'individu"; ce programme est "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 conséquentialistes.

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, tels 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, représenté par l'ajout du postulat P4. Qu'en est-il alors de la frontière entre ce type de rationalité et l'irrationalité ? Quelle place occupe chez Boudon les comportements irrationnels ? Il semble bien, *a priori*, que ce soit le domaine

des actions qui ne sont pas soutenus 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. A 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, italiques originales). Afin de tenter 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: 139-41).

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 rationalités alternatives existent, et qu'elles ont une influence sur les motivations des agents, mais pour l'instant, cette théorie n'est 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 semble désigner 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: ch. 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 se soit effectivement comporté de manière rationnelle (ou non), sans se soucier que l'agent 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 exhiber une "histoire causale" à laquelle l'agent puisse s'identifier[28] (Elster 1986a: 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 biens

[28] Cela constitue pour Elster une "idée préanalytique" (Elster 1989a: 6).

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 relative au choix qu'il a à effectuer[29]. 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, le conformisme (ou l'anti-conformisme), l'inertie, etc. En fait, sont considérés comme non autonomes les désirs face auxquels l'agent ne s'"identifie" pas pleinement, pour reprendre la citation plus haut.

Les critères de jugement et d'autonomie n'ont pas de définitions propres; ils sont plutôt compris comme des qualificatifs que l'on applique respectivement aux croyances et aux désirs, à la condition qu'ils n'exhibent pas certaines formes typiques d'irrationalité (Elster 1983: 24). Elster subdivise les comportements irrationnels en deux catégories. Les croyances et les désirs peuvent d'abord avoir été mal formés sans qu'aucune motivation non rationnelle ne soit intervenue ("cold irrationality"). Au niveau des croyances, il s'agit principalement des erreurs inférentielles. Un exemple de désir mal formé est le *framing*, qui survient lorsque les préférences sont influencées par un cadre cognitif particulier, en supposant qu'un autre cadre équivalent aurait généré un ordonnancement différent des préférences (Elster 1983: 25-26; 1989a: 23-24). La seconde catégorie comprend les irrationalités motivées ("hot"). Elle suppose que l'agent désire, la plupart du temps, agir de façon rationnelle; donc qu'il entretient la rationalité comme motivation. A 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 exhibe 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*,

[29] 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.

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).

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 1986a: 22-27; 1989a: 30-35). Plus tard (Elster 1999), il ne retiendra que les motivations émotive et rationnelle; la motivation à se conformer aux normes étant de nature surtout émotive, sans 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[30]. L'aspect le plus important, selon lui, de la motivation émotionnelle par rapport à la motivation rationnelle, c'est son côté "viscéral" et non conséquentialiste, au sens où l'influence émotionnelle pousse l'agent à dévaloriser les conséquences de ses gestes[31] (Elster 1999: 287). On ne peut simplement considérer les émotions comme un coût psychique (ou un bénéfice) venant modifier les valeurs d'utilité des alternatives, et ainsi, l'ordonnancement des préférences. Pour Elster, le caractère viscéral des émotions fait en sorte qu'en plus d'agir comme facteur d'évaluation des alternatives, 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

[30] 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).

[31] "Visceral arousal is an important criterion for deciding that a state is an emotion rather than a simple belief-desire complex" (Elster 1999: 247).

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[32].

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 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 occuring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences" (Elster 1998a: 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 il 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 raisons 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, pour les 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 (...)

[32] A 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.

[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, italiques originales). L'exemple qu'offre Boudon est frappant, car il s'agit, sans la nommer, de la position d'Elster. On pourrait qualifier cette position d'*irrationalité axiologique* : les valeurs sont essentiellement soutenues par les émotions[33], donc l'action motivée ainsi ne peut être pleinement rationnelle, alors que chez Boudon, les valeurs sont soutenues avant tout par de bonnes raisons, ou constituent en soi de bonnes raisons.

L'avantage principal, selon nous, de la théorie de 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 idéal-typiques, variant selon leur croyance en leurs habiletés scolaires (croyances cognitives) et leurs préférences pour un statut social élevé (croyances axiologiques). Il apparaît clairement, toutefois, que ces décisions de poursuivre ou non ses études sont de nature conséquentialistes : 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

[33] A ma connaissance, Elster n'affirme pas explicitement que les émotions fondent les valeurs; il s'agit ici d'une inférence à partir de ses discussions sur les normes sociales. Il serait selon moi difficile de réfuter l'affirmation de Etzioni à l'effet que des valeurs soutenues sans aucune émotivité ne serait que de vulgaires "tracts intellectuels" (Etzioni 1988: 105). Livet (2002) soutient que nos expressions émotives nous révèlent nos valeurs, une autre manière d'associer émotions et valeurs.

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 s'avérer sensible à la réputation familiale. Cette critique est invalide, car dans la théorie du choix rationnel, "égoïsme" signifie que l'utilité de l'agent lui est propre, mais 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, des ambiguïtés déterminantes demeurent. 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 le théorie du choix rationnel et certains "mécanismes sociaux" impliquant le non conséquentialisme.

1.4 Conclusion

La théorie du choix rationnel peut très bien s'accommoder des comportements non conséquentialistes. Elle peut faire "comme si" ils étaient conséquentialistes et ainsi les incorporer comme préférences, elle peut en conserver l'aspect non conséquentialiste et les modéliser comme contraintes à l'action, ou, si la modélisation s'avère trop ténue, elle peut en dernière instance les considérer comme non rationnels. Une critique qui implique que les individus ne sont pas toujours rationnels d'une manière instrumentale est sans objet dans le cadre de la théorie, car celle-ci modélise intentionnellement ses agents comme idéalement

rationnels, en leur attribuant des préférences *ad hoc* s'il le faut. Sa raison d'être n'est pas de décrire les individus tels qu'ils sont, mais de découvrir de quoi aurait l'air telle situation sociale, sous telles circonstances, avec des agents rationnels; et ensuite d'évaluer ses conclusions à la lumière des phénomènes sociaux réels. Nous pourrions faire le même constat à propos d'autres théories sociales, comme par exemple, la théorie de la démocratie délibérative. 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. Critiquer cette théorie sur la base que souvent, les individus se comportent de façon rationnelle et égoïste dans l'agora est également sans objet, car son objectif est d'élaborer des modèles de décisions collectives à partir de citoyens idéalement raisonnables, et ensuite de comparer ses résultats avec les procédures réelles. Selon nous, la théorie du choix rationnel ne fait pas fausse route en tentant de réifier les comportements non conséquentialistes; il s'agit seulement de se rendre compte qu'elle excelle dans certains domaines et moins dans d'autres, et qu'il n'y a pas qu'une seule théorie valide dans le champ des sciences sociales. Pour reprendre notre terminologie du début, la critique de la pertinence générale de la théorie du choix rationnel (ou encore, de la démocratie délibérative) se situe en amont, et sa critique circonstancielle, appelant à une substitution lorsque les modèles se révèlent insatisfaisants, se situe en aval de l'hypothèse de rationalité.

Nous avons choisi de traiter des valeurs et des normes comme représentants typiques du non conséquentialisme. Nous avons constaté qu'à l'intérieur des modèles de choix rationnel, nous n'avons en définitive d'autre choix que de modéliser ces comportements en termes de préférences ou de contraintes ordinaires, sous peine de violer les postulats de la théorie de l'utilité. Afin de prendre vraiment au sérieux les comportements non conséquentialistes, il nous faut sortir du cadre de la théorie du choix rationnel. C'est précisément ce qu'ont tenté d'accomplir Boudon et Elster, dans des optiques différentes. Sortir de la théorie du choix rationnel ne signifie pas s'en débarrasser, bien au contraire; on peut la conserver pour certaines circonstances, et se servir de théories alternatives pour d'autres circonstances, ou même tenter un "méta-modèle" qui userait de plusieurs modèles côte à côte. Sortir ainsi du choix rationnel signifie, par contre, que l'on doive abandonner, pour l'instant, la rigueur et la simplicité de ses modèles. En définitive, nous avons retenu la théorie des motivations mixtes de Elster, non pas

pour sa rigueur ou sa clarté, 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.

CHAPITRE II

LA DÉLIBÉRATION CIRCONSTANCIELLE EN THÉORIE DÉMOCRATIQUE

Nous avons proposé, dans le chapitre précédent, d'effectuer une distinction entre les motivations rationnelles et non rationnelles dans la détermination de l'action collective, car celles-ci auront des effets différents que, bien souvent, on ne peut négliger. Avant d'entreprendre l'élaboration de notre modèle d'action collective suivant ces grandes lignes, nous proposons dans ce chapitre de faire l'examen d'une théorie du champ politique, la démocratie délibérative, qui repose sur une structure analogue aux modèles sociaux que nous avons survolés précédemment, mais avec une différence fondamentale : bien qu'elle soit aussi de nature individualiste méthodologique, elle remplace l'hypothèse de rationalité par ce que nous pourrions appeler une hypothèse de "raisonnabilité", faisant de l'"agent" un "citoyen". Comme nous le verrons plus en détail, ce type d'individu se veut essentiellement *non rationnel*.

La démocratie délibérative a pour but d'élaborer les conditions de la pratique de la démocratie tel que les lois et les décisions produites par la délibération publique seront le plus juste possible pour la collectivité. C'est avant tout une théorie idéale, proposant comment les citoyens devraient se comporter afin d'en arriver à la société juste. Bien qu'il existe de nombreuses variantes de la démocratie délibérative, elles reposent toutes sur une conception "raisonnable" du citoyen : honnête, prédisposé à l'argumentation, ouvert aux opinions des autres et visant le bien-être de la collectivité; elles supposent aussi que ce citoyen ait la possibilité de participer à la délibération démocratique d'une manière libre et égale. Bien que son propos soit avant tout éthique, la démocratie délibérative ne peut éviter la modélisation de l'agrégation des voix des citoyens individuels en décisions collectives. Nous retrouvons donc nécessairement au sein de la démocratie délibérative des problèmes d'action collective, qui ne sont pas étrangers aux modèles plus rationnels de la décision politique. Notre hypothèse est que les modèles rencontrés en démocratie délibérative ne traitent pas de l'action collective de façon satisfaisante, au sens où ils règlent un peu trop rapidement des questions pourtant complexes telles que la recherche du consensus, la "force du meilleur argument", et d'autres

que nous aborderons tout au long de ce chapitre. Cette théorie, bien qu'elle offre des principes normatifs cohérents et certainement pas dénués d'intérêt, n'aborde pas adéquatement les questions d'application de ces principes, car, selon nous, elle s'attarde trop aux motivations éthiques non rationnelles et n'accorde pas assez d'importance aux motivations rationnelles.

Parallèlement, une autre théorie politique s'occupe précisément de la pratique actuelle de la démocratie : c'est la théorie du choix social, une théorie descriptive et explicative fondée sur le choix rationnel qui se penche sur les interactions entre individus rationnels (et souvent égoïstes) cherchant à maximiser leurs intérêts dans l'arène politique. Les adeptes de la démocratie délibérative ont fréquemment critiqué cette approche, observant que des individus purement rationnels, maximisant leur utilité dans l'arène politique, n'atteindront jamais la société juste. Nous proposons qu'une approche rationnelle "large", faisant place aux motivations non rationnelles préconisées par la démocratie délibérative, pourrait traiter des problèmes d'action collective de façon plus rigoureuse que la démocratie délibérative seule, sans pour autant mettre au rancart ses propositions éthiques. Après un survol de la démocratie délibérative axé sur le thème de la formation et de l'expression des préférences, nous allons aborder ce que nous nommerons la *délibération circonstancielle*, soit un ensemble de mécanismes sociaux ouvrant la voie à la possibilité de délibération parmi des citoyens rationnels. Nous allons donc chercher à voir comment de tels citoyens pourraient en arriver à adopter les comportements prescrits par la démocratie délibérative. La délibération circonstancielle interpelle les démocrates délibératifs à deux niveaux : d'abord ses mécanismes sociaux leur fournissent des outils théoriques utiles à la modélisation des interactions sociales et de plus, la démonstration que la délibération rationnelle est possible devrait enrichir cette théorie et ouvrir la voie à des variantes plus réalistes. Avant d'aborder les mécanismes de la délibération circonstancielle, nous allons nous pencher sur les difficultés rencontrées par la démocratie délibérative dans l'application de ses principes normatifs.

2.1 La pratique délibérative

Dans cette section nous nous intéresserons aux comportements des citoyens à l'intérieur des modèles politiques préconisés par diverses variantes de la démocratie délibérative. Cet examen de la pratique délibérative nous permettra d'évaluer comment ces

théories effectuent le passage crucial de la prescription normative à l'application concrète. Il ne s'agit pas d'une étude empirique de la délibération, mais bien d'une modélisation abstraite de l'agora fondée sur certaines conceptions du citoyen individuel et des procédures démocratiques. Nous verrons comment certains principes fondamentaux de la démocratie délibérative tels l'argumentation raisonnable, l'égalité, la réciprocité, et d'autres prennent vie dans ces modèles. L'étude de l'application de ces principes normatifs peut prendre de multiples formes. Étant donné que nous avons choisi d'aborder ces modèles délibératifs d'un point de vue rationnel, nous nous concentrerons sur deux aspects chers à la théorie du choix rationnel, soit la *formation* et l'*expression* des préférences individuelles. Cette discussion tournera autour du concept d'*autonomie* des préférences, un concept relié à la liberté individuelle préconisée par la démocratie délibérative. Nous débuterons par la question de la formation autonome des préférences : le consensus s'atteint-il par réflexion libre et rationnelle ou par conformisme à l'opinion majoritaire ? Nous passerons ensuite à la question de leur expression autonome : jusqu'à quel point l'exigence d'égalité contraint-elle les arguments que le citoyen peut avancer ? Nous verrons que sur ces questions, certaines variantes importantes de la démocratie délibérative exhibent des contradictions théoriques ainsi que des effets pervers de nature psycho-sociologique.

La formation des préférences

Dans la théorie du choix rationnel, l'action est déterminée par les désirs et les croyances de l'agent. Selon le modèle de Jon Elster que nous allons utiliser tout au long de cet article, pour qu'une action soit substantiellement rationnelle les désirs doivent être autonomes et la formation des croyances doit exhiber un certain jugement. De plus, les liens entre désirs et croyances sont sujets à évaluation. Commençons par l'autonomie des désirs. Les phénomènes les plus courants d'irrationalité dans la sphère politique concernent les comportements conformistes (ou anticonformistes), le respect irréfléchi des traditions ou la révolte contre celles-ci, etc. Aux deux extrêmes, les préférences "adaptatives", soit la correspondance parfaite avec la situation actuelle, perpétuent le *statu quo* social alors que les préférences "contre-adaptatives", illustrées par le dicton voulant que l'herbe toujours plus verte sur le terrain du voisin, aboutissent à un choix social impossible à mettre en oeuvre (Elster 1986b: 109-110) ou

du moins très instable[34]. Ces phénomènes sont constitutifs d'une réalité indéniable concernant la formation de n'importe quelle préférence, soit l'influence de la société par l'éducation et l'habitude. Aucun être humain ne peut se prétendre complètement immunisé des valeurs de la société dans laquelle il vit. Les individus auraient donc un penchant naturel pour la perpétuation des valeurs et des institutions existantes. Par exemple, l'adoption du vote proportionnel est beaucoup plus difficile à obtenir dans un pays avec une longue tradition de démocratie directe par circonscriptions que dans un pays sans véritable tradition démocratique. Bien que ne constituant pas une irrationalité au sens fort, cet effet de préservation d'acquis (*endowment effect*, Sunstein 1993: 199) prouve que les individus ne forment pas leurs préférences en toute liberté.

Débutons par la formation des préférences en démocratie délibérative. Par censure de groupe[35] on entend les divers modes de pression sociale contre les prises de positions ouvertement égoïstes. La délibération se fie notamment à la "force civilisatrice de l'hypocrisie" pour convertir les individus égoïstes. Postulant l'existence d'une norme sociale contre l'expression de préférences égoïstes en public, ce mécanisme stipule qu'un individu égoïste, à force de s'exprimer d'une façon faussement impartiale, en viendra à sincèrement adopter ces principes (Elster 1994a: 190). Bien que justifiables en termes de justice sociale, de telles préférences ne répondent pas adéquatement au critère rationnel d'autonomie. Si le mécanisme opérationnel est la réduction de dissonances, il devient difficile de croire que de telles préférences ont été intentionnellement formées (Elster 1986b: 113; Johnson 1998: 172). Difficile donc de faire la part des choses entre préférences intentionnelles ou non, entre préférences "réelles" ou stratégiques. Aussi, la distinction entre intérêts privés et publics semble trop simpliste (Elster 1986b: 119). Par exemple, lorsqu'un groupe exclu de l'agora

---

[34] Dans la logique des ensembles, avec l'ensemble des préférences P d'un individu et l'ensemble faisable F nous avons l'adaptation lorsque $P \cap F = P$ et la contre-adaptation lorsque $P \cap F = \varnothing$. Dans sa définition de l'autonomie, Elster demande que P et F s'entrecoupent mais pas entièrement. L'instabilité de l'implémentation sociale des préférences contre-adaptatives s'illustre de la façon suivante : lorsque le gouvernement implémente des préférences formées de telle manière, il se trouve à modifier F afin d'y inclure P mais alors les citoyens contre-adaptatifs (irrationnels) vont rejeter la nouvelle donne en exprimant de nouvelles exigences au-delà de F.

[35] L'emploi du terme "censure" tout au long de ce chapitre n'implique pas de connotations négatives, ni un jugement de valeur envers la démocratie délibérative. Toute théorie politique encadre d'une certaine manière les faits et gestes des citoyens, c'est ce que le terme cherche à saisir.

exige sa place dans l'espace public, il le fait au nom de ses intérêts qui, on le suppose, ne sont pas représentés adéquatement (Johnson 1998: 174; Knight, Johnson 1994: 288).

La persuasion, tout comme la censure, se bute aussi à l'autonomie des préférences. Les théoriciens délibératifs insistent sur une persuasion "rationnelle" fondée sur un débat raisonnable et un ajustement volontaire et justifié par l'individu de ses positions initiales. Toutefois, il n'existe aucun moyen sûr de distinguer entre une telle forme de persuasion et un simple mimétisme ou un effet d'entraînement de la masse (Elster 1986b: 116-17), les deux dernières formes étant manifestement irrationnelles. La persuasion peut aussi induire une nouvelle inégalité substantielle car certains individus sont plus versés que d'autres dans l'art oratoire et disposent donc d'un avantage sur les autres. Bien que certains aient tenté de critiquer cette inégalité (Young 1996), il demeure extrêmement difficile de voir comment la force persuasive pourrait être uniformisée préalablement à la discussion. La démocratie délibérative ne peut espérer une transformation massive des préférences par la simple discussion publique (Johnson 1998: 174).

L'uniformisation des préférences n'étant pas moralement désirable si elle implique la violation de l'autonomie individuelle[36], nous sommes alors confrontés à une pluralité des conceptions du social, ce que Rawls appelle les doctrines compréhensives. Le conflit doctrinaire prend une toute autre dimension que la simple divergence d'intérêts; il met en présence des valeurs fondamentales, telles les croyances religieuses, qui ne peuvent être objets de négociations. Tout ce que l'on peut espérer pour éviter le conflit, c'est la tolérance. Chez Rawls, le citoyen idéal est qualifié de "raisonnable", un concept assez près de celui rencontré habituellement en démocratie délibérative, incluant en particulier la tolérance et l'impartialité. Les valeurs sociales de tels citoyens se nomment, cela va de soi, doctrines compréhensives raisonnables. Le conflit doctrinaire déraisonnable ne devrait donc pas survenir dans un système délibératif bien ordonné car, bien sûr, les doctrines déraisonnables y sont éliminées dès le départ (Rawls 1995, ch. II).

---

[36] Ou même un doute lancinant sur cette autonomie, car dans bien des cas il nous sera impossible de faire la part des choses.

Malheureusement, la réalité n'est pas si facile et les conflits impliquant des doctrines déraisonnables y seront bien souvent insolubles. D'abord, ce qui est raisonnable chez Rawls constitue en soi une doctrine particulière (Johnson 1998: 168-69). Même si Rawls prétend que ses principes de justice servent de fondation aux doctrines raisonnables (Rawls 1995: 183), cela ne fait que repousser le problème à un niveau supérieur. Ancrer ses préférences exprimées à une doctrine comporte un avantage de crédibilité, car on abandonne avec beaucoup plus de peine une valeur morale qu'une valeur d'utilité. Cela peut aussi servir la communauté en forçant un débat de fond sur les problèmes sociaux plutôt qu'une lutte d'intérêts superficiels. Mais son principal effet pervers sera de durcir les positions jusqu'à l'impasse politique, surtout si l'engagement préalable se prend en public: "(...) les mêmes motifs qui font que les interlocuteurs ont adopté d'emblée une attitude d'argumentation plutôt qu'une attitude de négociation les incitera également à tenir leurs engagements" et à réduire les chances de succès de la délibération (Elster 1994a: 248). Comme nous le remarquons fréquemment dans nos assemblées politiques, la publicité des débats amène les différents groupes à vouloir se démarquer idéologiquement; ils tendent ainsi vers la rhétorique (Elster 1986b: 118).

Gutmann et Thompson répondent à ce problème en suggérant le principe délibératif de réciprocité. Face à un désaccord moral fondamental, les citoyens mus par l'idéal de réciprocité abandonnent leurs intérêts personnels et, autant que possible, leurs arguments moraux incompatibles avec la position morale opposée. Sans nécessairement résoudre le conflit, ils atteignent à tout le moins un certain niveau de respect mutuel (Gutmann, Thompson 1996, ch. 2). Ces théoriciens délibératifs admettent que les conflits peuvent se durcir et même augmenter en nombre sous ces conditions, mais tout devrait relativement bien se passer en postulant des citoyens de bonne foi. La lacune fondamentale d'une telle proposition, c'est qu'en politique la bonne foi des participants résout énormément de conflits[37]. Évidemment ce n'est pas là un postulat suffisant, même s'il est longuement décortiqué avec le principe d'"accommodement moral" (Gutmann, Thompson 1996: 79-85)[38]. Nous retrouvons aussi cette bonne foi dans ce

[37] Cohen semble partager cette foi : "The structure of discussion, aimed at solving problems rather than pressuring the state for solutions, would encourage people to find terms to which others can agree. And that would plausibly drive argument and proposed actions in directions that respect and advance more general interests" (Cohen 1996: 113).

[38] Nous retrouvons dans ce concept des prescriptions de comportement comme la cohérence de ses principes moraux, la reconnaissance de l'autre, l'ouverture d'esprit ou la minimisation des points

court passage : "Moral argument can arouse moral fanatics, but it also combats their claims on their own terms" (Gutmann, Thompson 1995: 106). On voit mal comment on pourrait "combattre" les fanatiques au niveau argumentatif sans supposer qu'ils soient eux-mêmes ouverts au débat, ce qui est une contradiction. Cette théorie doit spécifier une autre méthode plus convaincante de négociation avec les fanatiques ou même tout citoyen qui ne discute pas pleinement de façon réciproque. Bien que la volonté des démocrates délibératifs sur la question des conflits politiques soit de favoriser la tolérance et le respect mutuel - objectifs parfaitement louables en soi -, force nous est de constater que ce type de délibération peut augmenter et rendre plus insolubles les conflits; en cela il n'est pas certain que la solution délibérative soit la meilleure.

L'expression des préférences

Autant la théorie du choix social que la théorie délibérative élaborent minimalement des *procédures* justes d'élaboration de politiques gouvernant la société. Toutes deux se fondent sur un individualisme méthodologique. Du côté épistémologique, elles dotent l'individu de certaines motivations intentionnelles encadrées par des critères fondamentaux de justice sociale. Les politiques adoptées trouvent leur justification dans la procédure les ayant engendré. La raison d'être des théories est la même: les citoyens ont le pouvoir de forger de façon libre et intentionnelle une société viable, au double sens de capacité de survie de la société et de bonne vie pour ses membres. Les deux théories veulent absolument se distinguer des conceptions compréhensives du politique où le "philosophe-roi" détermine lui-même les institutions sociales en faisant fi des préférences individuelles de ses sujets, car évidemment cette méthode est foncièrement antidémocratique. Toutefois, ces théories vont quand même instaurer des limites aux arguments pouvant être présentés en public. Pour ce qui est du choix social, ces limites se veulent minimales; exception faite des arguments violant l'intégrité et la dignité de la personne, la plupart des arguments seront admis, même ceux fondés sur la position de pouvoir du locuteur. Il n'en est pas de même en démocratie délibérative, qui se veut beaucoup plus contraignante à ce sujet.

---

de désaccord. Bien entendu, adopter une attitude morale de résolution de conflits va réduire les conflits et nous aider à vivre avec le résiduel insoluble potentiel, mais au-delà de cette vérité plate, Gutmann et Thompson ne nous en disent pas beaucoup plus.

Amorçons cette discussion avec la définition de Rawls de la justice procédurale. Il nous offre quatre types. D'abord, la justice procédurale *parfaite* propose un critère d'évaluation des fins indépendant de la procédure, ainsi qu'une procédure telle que des individus rationnels atteindront de plein gré le résultat escompté. Il va sans dire que cette coïncidence entre justice et intérêt particulier est plutôt inusitée; Rawls propose donc la justice procédurale *imparfaite* où la procédure peut ne pas garantir la convergence vers le résultat demandé par le critère de justice substantielle. Dans les deux cas, le résultat correct est déterminé à l'avance, la question devient de savoir comment l'atteindre démocratiquement. En justice procédurale *pure*, le résultat est indéterminé et trouvera sa justification dans la justesse de la procédure suivie. Rawls spécifie que ce type de justice doit être complété par des institutions justes, sinon on risque le chaos; selon lui, celles-ci sont choisies dans la position originelle, alors que le choix social impose des institutions libérales[39]. Une justification supplémentaire apparaît en justice procédurale *quasi-pure*, soit le spectateur impartial respectant certains critères substantiels de justice. Le résultat de la procédure est ainsi comparé à ce que des citoyens hypothétiques obéissant à une justice substantielle auraient choisi. Ce résultat hypothétique ne se veut pas unique, mais couvre un certain éventail de résultats justes (Rawls 1971: 85-87, 362). C'est une façon originale de réintroduire le substantiel dans un principe qui avait pour but d'en limiter la portée.

Plusieurs démocrates délibératifs vont employer cette stratégie de la double légitimation consistant à imposer un critère supplémentaire de légitimation du résultat, au-delà du respect de la procédure juste. Pour Gutmann et Thompson, la délibération doit se dérouler en contexte de respect mutuel et de réciprocité, soit l'attention portée aux arguments d'autrui, la disposition à réviser ses propres arguments, etc. C'est une forme de justice substantielle présente dans la procédure même. Le résultat n'est légitime que si les citoyens ont respecté les critères de réciprocité (Gutmann, Thompson 1995: 104-6). La procédure juste ne suffit pas : "From a deliberative perspective, the problem with relying on bargaining as a substitute for

[39] Pour Gutmann et Thompson (1995: 99), on retrouve autant dans les fondements substantiels de Rawls ("justice constitutionnelle") que dans ceux du choix social ("justice procédurale") des principes nécessaires à la procédure juste (droit de parole, etc.) et des principes extra-procéduraux nécessaires à la survie de la société (revenu minimum, etc.). Ces principes servent de contraintes à la procédure. Les deux diffèrent lorsque Rawls propose en plus des principes non nécessaires à la société et non contraignants.

moral reasoning, *even within political institutions that are fully just*, is that it rests on too thin a conception of what citizens owe one another in an increasingly interdependent society" (Gutmann, Thompson 1996: 58, italiques rajoutées). Cohen propose une version légèrement différente : si deux procédures donnent exactement le même résultat, on doit préférer celle qui laisse le plus de place à la délibération, "because of the greater confidence in the deliberative character of the process and the increased confidence in the outcomes that results" (Cohen 1986: 37). On remarquera qu'il est ici question de confiance et non de légitimité, mais c'est passablement la même chose.

Gutmann et Thompson adoptent également une forme de justice procédurale quasi-pure. Ils nous offrent un exemple de délibération se déroulant dans un train : les passagers doivent décider s'ils ont le droit de fumer à l'intérieur du wagon. Quelques uns (une minorité) invoquent l'argument de la détérioration de la santé de tous, malgré cela le vote est pris et la permission de fumer l'emporte. Les auteurs voient là une situation où la procédure majoritaire perd de sa légitimité (Gutmann, Thompson 1995: 94-95). Cette procédure constitue pour eux une théorie politique dite de "premier ordre", c'est-à-dire qu'elle s'impose en rejetant les théories alternatives. Nous faisons donc face à un désaccord entre deux visions de la politique : la procédure majoritaire et la défense du bien-être collectif proposée par la minorité. Aucune des deux ne doit l'emporter sur l'autre *a priori*. La démocratie délibérative est, quant à elle, une théorie de "second ordre" se situant au-dessus des autres théories, dans une position permettant de les juger (Gutmann, Thompson 2000). En fait, cela ressemble plus à l'argument du spectateur impartial de Rawls jugeant des résultats selon certains principes. Mais ils s'aventurent plus loin. Selon eux, la démocratie délibérative elle-même n'échappe pas à son auto-critique; elle fait donc l'objet de ses propres principes. Bien que ce critère préserve la cohérence de la théorie, il la vide également de son contenu. Tous les principes de la démocratie délibérative se veulent provisoires, et il ne nous reste en définitive que le principe fondamental de réciprocité (qui peut être lui-même rejeté, mais alors ce sera la fin de la démocratie délibérative). Cette technique permet à Gutmann et Thompson d'introduire des

principes de justice substantielle sans vraiment avoir à les inclure formellement dans la théorie[40].

Une troisième variante de la double légitimation est la "validation récursive" de Seyla Benhabib. Partant d'une définition délibérative classique de la procédure juste - délibération libre et raisonnée, égalité morale et politique des citoyens -, elle n'impose aucune restriction aux arguments exprimables dans l'agora et aux résultats conséquents : "Procedures can neither dictate outcomes nor define the quality of the reasons advanced in argumentation nor control the quality of the reasoning and rules of logic and inference used by participants" (Benhabib 1996: 72). Ceci représente le premier niveau de délibération. La seconde légitimation fait son apparition au "méta-niveau", une délibération permettant de juger des résultats du niveau précédent. Cette validation récursive permet aux adversaires de la solution démocratiquement proposée de s'y opposer et ainsi de forcer l'élaboration d'une autre solution : "(...) only the *freely given assent of all concerned* can count as a condition of having reached agreement in the discourse situation" (Benhabib 1996: 79, italiques dans l'original).

La double légitimation s'adresse à une difficulté bien connue en démocratie délibérative, la tension entre le citoyen "idéal" raisonnable et le citoyen "réel" dont les actes peuvent dévier de l'idéal de justice. Par souci de réalisme, les partisans d'une forme ou d'une autre de la double légitimation ne veulent pas imposer l'idéalisme au niveau du citoyen mais en fait, ils ne font que repousser cet idéalisme à un niveau supérieur, ce qui n'est guère mieux. Le besoin de deux mécanismes de légitimité dans ces théories doit également être compris à l'intérieur d'une tension entre l'intuition démocratique de laisser les citoyens décider par eux-mêmes dans le cadre d'une procédure juste et l'intuition paternaliste de juger le résultat selon des critères substantiels extra-procéduraux. Bien qu'ancrée d'une certaine manière dans la procédure, la justice quasi-pure, la réciprocité et la validation récursive émanent d'une justice substantielle, sinon pourquoi ne pas choisir le tirage au sort comme processus décisionnel ? Si le tirage s'effectue dans des conditions équitables, voilà une procédure tout à fait juste. Le fait

[40] Le problème épistémologique qui se pose, c'est que les principes de *toute* théorie, dans *tous* les domaines de la science, sont essentiellement provisoires. Même la théorie politique la plus totalitaire et moralement hermétique finit par transformer ses principes au fil du temps. Donc, cette définition particulière de la démocratie délibérative n'apporte rien de concret et sombre dans un profond relativisme.

de préférer la délibération au tirage démontre que l'on recherche l'imposition d'un *certain type* de procédure, pas seulement une procédure quelconque respectant des critères d'impartialité et d'équité (Cooke 2000: 950-52). On attribue souvent à une telle procédure une valeur civilisatrice propre. Les individus "raisonnables", en écoutant les autres et en évaluant leurs points de vues sans préjugés, en viennent à devenir de meilleurs citoyens. La procédure prend la forme d'une fin, pas seulement d'un moyen. Cette fin est nécessairement substantielle.

L'égalité des citoyens constitue une part importante de la justice substantielle délibérative. Cela peut prendre plusieurs formes. Il y a d'abord l'égalité politique, le fondement de toute théorie démocratique, soit le droit de vote égal (une personne, un vote), le droit égal de se présenter aux élections, etc. Vient ensuite l'égalité morale, c'est à dire l'absence de toute autorité morale formelle dans le débat. Les participants n'ont pas à se plier aux arguments d'autrui seulement à cause de sa supériorité comme membre du clergé, philosophe, etc. Benhabib (1996: 68), par exemple, fonde la démocratie délibérative sur ce qu'elle nomme le "modèle discursif de l'éthique", comprenant l'égalité de participation et de propositions d'arguments et le droit égal de remettre en question les sujets à l'ordre du jour ainsi que les règles délibératives. Rawls maintient à peu près le même propos avec son principe de participation égale (Rawls 1971: 221-34). Il n'est pas question dans ce principe de redistributions économiques ou sociales[41].

D'autres théoriciens, par contre, vont exiger une égalité plus substantielle. La justification générale est simple : afin d'atteindre une authentique égalité politique, il faut éliminer les facteurs externes conférant un avantage injuste à certains participants, comme la position sociale ou la propriété de ressources économiques. Nous retrouvons de tels principes chez Cohen : "The participants are substantively equal in that the existing distribution of power and resources does not shape their chances to contribute to deliberation, nor does that

[41] Il est à noter que chez Rawls le principe de participation égale se limite à des questions politiques essentielles comme la formation de la Constitution. Les exigences égalitaires sont beaucoup plus fortes dans la position originelle, mais ceci se veut un exercice abstrait de découverte des principes de justice et non une définition des individus réels dans des discussions concrètes. Rawls demande seulement que les citoyens exhibent en général une *disposition* morale compatible avec la position originelle, peu importe s'ils l'appliquent ou non (Rawls 1971: 505). Bien sûr, l'application conduira à une société plus juste, mais ce n'est pas une exigence formelle.

distribution play an authoritative role in their deliberation" (Cohen 1989: 23). Ses propos semblent signaler une forme d'auto-censure; on laisse ses avantages au vestiaire en entrant dans l'agora. Lorsque les démocrates délibératifs posent l'égalité substantielle en termes d'un idéal de citoyenneté, en érigeant des barrières théoriques à l'expression des préférences, ils ne nous informent aucunement sur les mécanismes permettant d'y arriver. Rawls a bien compris ce problème : "(...) the principle of participation applies to institutions. It does not define an ideal of citizenship; nor does it lay down a duty requiring all to take an active part in political affairs" (Rawls 1971: 227). Pour lui, l'égalité individuelle substantielle est une question de sociologie politique, pas d'une théorie de la justice (Rawls 1971: 226-27). Lorsque Cohen (1989: 30-32; 1996: 108-10) exige une certaine redistribution égalitaire sous forme de financement public des partis, il penche du côté de Rawls mais sa volonté de situer l'égalité au niveau du citoyen revient constamment.

La variante la plus répandue de l'égalité substantielle se retrouve dans le principe délibératif habermasien de la "force du meilleur argument". Les arguments devraient être évalués selon leur vérité, leur sincérité et leur valeur morale (souvent l'impartialité), et non selon l'avantage social ou matériel de celui ou celle qui les propose. Ce principe agit comme une forme de censure à la discussion, régulant ce qui peut ou ne peut être avancé comme arguments. Une proposition peut être rejetée si elle n'est pas fondée sur des "raisons acceptables" (Cohen 1989: 22). La démocratie délibérative ne s'intéresse pas aux gens n'appliquant pas le principe de réciprocité (Gutmann, Thompson 1996: 55). L'État peut même intervenir afin de censurer leurs opinions (Rawls 1995: 91). Cohen allait encore plus loin dans sa théorie du populisme épistémique qu'il adoptait il y a plusieurs années (il a depuis changé d'opinion). Les citoyens ne doivent pas exprimer leurs préférences, mais se limiter plutôt à évaluer les propositions à l'étude selon les critères d'une "volonté générale" correspondant plus ou moins aux principes rawlsiens de justice. Il est permis de modifier la procédure si jamais la tentation de recourir aux préférences s'avère trop forte (Cohen 1986: 34-37). Il faut toutefois admettre que la démocratie délibérative n'endosse pas généralement ce genre de proposition.

Outre le questionnement sur les conséquences d'un tel contrôle sur la démocratie, il existe une autre critique plus axée sur la théorie. Elle est amenée un peu indirectement par

Young (1996) dans son exposé post-moderniste sur la validité de la "force du meilleur argument". Elle voit dans l'égalisation des participants un danger d'uniformisation, voire de conservatisme. Ainsi, la discussion raisonnable exclurait les voix plus émotives ou encore provenant de cultures ne pratiquant pas le débat "raisonnable" au sens des démocrates délibératifs[42] (Young 1996: 122-24). Elle s'en prend dans son texte au caractère mâle et Blanc des institutions délibératives proposées par les théories, mais nous pouvons en déduire un autre problème, celui de la relativité des prémisses d'égalité délibérative. En effet, pourquoi s'arrêter à l'égalité des positions sociales ou des ressources ? Certains possèdent plus de talent persuasif que d'autres, certains sont davantage attirés vers la compétition intellectuelle, etc. Pour Young, la diversité est une ressource nous permettant de découvrir les besoins réels des autres, pas une variable à niveler (Young 1996: 126-28). Elle nous amène à réfléchir sur la pertinence de l'argumentation d'égal à égal en démocratie délibérative[43]. Elle soulève également un coin du voile recouvrant la complexité d'une conception adéquate de l'égalité.

Nous avons vu dans cette section que la sphère politique affecte la formation autonome des préférences, par les effets de préservation d'acquis, d'entraînement de la masse et de l'adaptation au possible. Cohen (1989: 25-26) a relevé ce problème potentiel en politique, mais selon lui les principes délibératifs, notamment le "pouvoir de la raison", assurent l'autonomie des préférences. Il n'y aurait donc pas de problèmes, malheureusement il ne précise pas vraiment comment on en arrive au juste à cette autonomie. L'introduction de considérations rationnelles nous apprend qu'il est impossible de distinguer clairement, lors d'un débat public, entre une préférence authentiquement autonome et un simple effet de conformisme. Le "pouvoir de la raison" ne garantit donc rien. Plus encore, la délibération à grande échelle a tendance à exacerber les mécanismes contrant la formation autonome des préférences (Elster 1998b: 107-9), c'est le phénomène bien connu des effets de foule. Des représentants élus ou des leaders émergeants opportunistes peuvent se servir des effets de foule pour assouvir leurs

[42] Johnson (1998: 166) abonde dans le même sens lorsqu'il observe qu'à prime abord, la désobéissance civile, par son caractère agressif et impertinent, n'a pas sa place en démocratie délibérative même si ses demandes sont parfaitement légitimes.

[43] Young ne rejette pas vraiment la démocratie délibérative, elle la remplace par la "démocratie communicative" où des individus affirmant leurs différences en viennent à transformer leurs préférences par un processus d'apprentissage et de respect de l'autre. Benhabib (1996: 82) se demande avec justesse quelle est la différence entre ce mécanisme et le consensus raisonnable que l'on rencontre habituellement en démocratie délibérative.

propres fins, par la démagogie et la rhétorique. Afin de se prémunir de ces tendances, il est parfois préférable pour les représentants de se réunir à huis clos. Cette violation du principe de publicité peut être justifiée de façon délibérative lorsque les citoyens décident librement (par vote ou consensus) qu'il serait mieux au nom de la collectivité que les représentants ne soient pas tentés de sombrer dans la démagogie (Luban 1996: 189-92; Gutmann, Thompson 2000: 176-77). En effet, la délibération publique et transparente sur des sujets controversés ne se fait généralement pas à tête reposée.

Les concepts de réciprocité et de doctrine compréhensive raisonnable ont comme but de s'accorder avec le "fait du pluralisme", soit la reconnaissance que nous n'arriverons jamais à des choix unanimes de société et que de toute façon il serait moralement indésirable d'en arriver là. Toutefois, ces mêmes principes peuvent conduire à un durcissement idéologique, réduisant ainsi considérablement la portée de la délibération démocratique. Gutmann et Thompson ont reconnu ce problème mais comme nous l'avons vu, ils ne nous offrent pas de véritables méthodes de résolution. Les différentes variantes de "double légitimité" proposent des mécanismes d'agrégation des préférences individuelles assurant le respect d'une certaine justice substantielle, mais si la démocratie délibérative veut s'engager sur ce terrain extrêmement fertile, il faudrait qu'elle nous offre des mécanismes beaucoup plus élaborés que la "validité récursive" ou son statut de théorie de second ordre. Enfin, nous avons constaté que l'égalité des citoyens dans l'agora est loin d'être un concept simple, car il existe une multitude d'implémentations possibles des mêmes critères généraux d'égalité.

2.2 La délibération circonstancielle

Ayant entrevu les conséquences de l'introduction de certains principes rationnels dans la modélisation de l'agora, nous allons maintenant nous pencher plus sérieusement sur les manières dont la théorie du choix rationnel pourrait nous venir en aide concernant les questions d'implémentation des principes normatifs de la démocratie délibérative. D'abord, une précision s'impose. La théorie du choix rationnel n'exige aucunement que l'agent se comporte de façon égoïste. On rencontre souvent cette hypothèse car ce sont les économistes qui se servent le plus du choix rationnel et pour eux, l'agent égoïste génère de meilleurs modèles. Mais dans un domaine comme la politique, on peut appliquer le choix rationnel sans présupposer le caractère

égoïste ou altruiste du citoyen. Il est donc possible de considérer en choix social, par exemple, des motivations individuelles inspirées des idéaux délibératifs. En fait, on pourrait affirmer grossièrement qu'un système politique fondé sur le choix social et ne comprenant que des participants motivés par l'éthique délibérative serait identique à ce que la démocratie délibérative elle-même propose. Bien sûr, la leçon fondamentale que nous enseigne le choix social est que nous ne pouvons compter sur une telle communion morale. La possibilité de délibération parmi des citoyens rationnels nous permettra de résoudre certaines des difficultés rencontrées plus haut en démocratie délibérative, mais non sans en faire apparaître d'autres. Nous allons d'abord nous intéresser aux motivations individuelles de type délibératif pour ensuite traiter des possibilités d'action stratégique entourant les choix collectifs, le tout dans une perspective rationaliste.

Les motivations individuelles

Les démocrates délibératifs ont fréquemment tendance à opposer le "raisonnable" au "rationnel", menant ainsi au conflit quasi-perpétuel entre démocratie délibérative et choix social[44]. Rawls considère notamment la possibilité d'individus rationnels altruistes, mais en tant qu'individus exprimant des intérêts personnels en tenant compte du bien-être d'autrui. Il poursuit : "Ce qui manque aux agents rationnels, c'est la forme particulière de sensibilité morale qui sous-tend le désir de s'engager dans une coopération équitable comme telle (...)"[45] (Rawls 1995: 79). Nous allons tenter de démontrer que cette "sensibilité morale" peut faire partie des motivations de l'agent rationnel et qu'elle peut être provoquée ou facilitée par certains mécanismes socio-psychologiques. Pour ce faire nous allons surtout nous baser sur le modèle d'Elster, qui nous invite à considérer les normes sociales et les émotions comme sources supplémentaires de motivations chez l'agent rationnel.

---

[44] Le choix social n'est certainement pas innocent dans ce conflit, ainsi le plaidoyer de Sen (1986) pour que ses théoriciens respectent les choix éthiques.

[45] Gutmann et Thompson formulent la même critique de la rationalité : "Even if citizens were to bargain under conditions of approximate equality, the results might still fail to meet the minimal standards of sociability that a reciprocal perspective would specify" (Gutmann, Thompson 1996: 58).

Il existe chez Elster trois catégories de motivations, soit l'intérêt, la raison et la passion[46]. Sa définition de la raison se rapproche beaucoup de celle d'Habermas : l'agent s'engage à respecter les principes de vérité propositionnelle, de justesse normative et de sincérité (Elster 1999: 337). Le mécanisme fondamental à l'oeuvre ici est que dans tout débat public, les participants sont aux prises avec une norme sociale favorisant l'argumentation fondée sur la raison. Pour Elster il s'agit plus que d'une simple norme, ce type de comportement se veut quasiment une "vérité conceptuelle" (Elster 1999: 372). Il existe une relation fondamentale, inaliénable entre le débat public et la raison : "To say, in a public debate, 'We should choose policy A because it is good for me', is to show a fundamental lack of understanding of what it *means* to offer an argument for something" (Elster 1999: 373, italiques originales). Toutefois, nous pouvons nous en tenir à la norme sociale sans que cela n'inquiète notre propos, car la frontière entre une norme sociale forte et une "quasi-vérité" est plutôt ténue; Elster l'admet également.

Résumons très brièvement la conception d'Elster des normes sociales. Celles-ci se fondent sur deux émotions, prenant ainsi deux formes distinctes : la honte, une auto-évaluation négative de sa personne et la culpabilité, une auto-évaluation négative d'un geste commis. La honte est provoquée par le mépris d'autrui suite à la violation d'une norme sociale forte, alors que la culpabilité émane du jugement négatif, provenant de soi ou d'autrui, d'un geste répréhensible, impliquant une norme sociale de moindre importance (Elster 1999: 145-56). Ces émotions négatives nous portent à redéfinir nos actions de manière à éviter la douleur psychologique associée. Elles peuvent induire la transmutation d'une motivation (intérêt, raison ou passion) en une autre, ou encore sa fausse représentation (*misrepresentation*) (Elster 1999: 332). La transmutation représente un conflit entre le désir de promouvoir son intérêt personnel et celui de maintenir une certaine estime de soi (*positive self-image*) (Elster 1999: 369). C'est ici que l'on retrouve la possibilité de "sensibilité morale" de Rawls, si par "image positive" on considère le comportement raisonnable - au sens délibératif - dans l'agora[47]. La fausse

[46] Comme il s'agit d'une variante du choix rationnel, il va de soi que l'intérêt jouit d'une primauté théorique.

[47] Pour que cette proposition soit vraiment solide, il faudrait démontrer empiriquement qu'un bon nombre de citoyens jouissent de ce type d'idéal personnel. Nous ne nous engagerons pas ici dans un tel travail, toutefois nous pouvons aisément constater qu'une telle disposition existe bel et bien, et d'une manière assez fréquente.

représentation se veut une variante hypocrite de la transmutation, les motivations n'y sont pas transformées au sens fort où la motivation originale cède sa place à une autre, elle est ici simplement masquée, maquillée afin de bien paraître au yeux d'autrui[48]. Dans bien des cas, nous serons à des lieues de la "sensibilité morale", mais comme nous le verrons, il existe certaines classes de situations où la fausse représentation peut se montrer compatible avec les propositions de la démocratie délibérative.

Ce qui nous intéresse pour notre propos, ce sont les transmutations et les fausses représentations de l'intérêt en raison. En soi, la transmutation présente des similitudes avec le modèle du citoyen que propose la démocratie délibérative, soit l'agent doté d'intérêts privés qui, au sein de la procédure délibérative, en vient à exprimer des propositions de nature impartiale et sensibles aux demandes d'autrui. Alors que la démocratie délibérative fonde le raffinement des préférences sur l'expression simple d'une certaine moralité politique *a priori* ainsi que sur les effets bénéfiques de la participation politique et du débat raisonnable, la stratégie de la transmutation cherche à localiser ce raffinement dans le profil psychologique de l'agent. Celle-ci aussi accepte l'idée de moralité *a priori*, mais précise de plus que la préférence exprimée par l'agent est le résultat d'un certain conflit interne entre cette moralité et ses intérêts privés. Les implications de ce niveau supplémentaire de formation de préférences sont attrayantes.

L'impartialité constitue la forme la plus commune de motivation orientée vers le collectif. Toutes les variantes de la démocratie délibérative l'exigent de ses citoyens idéaux. Pour Elster, l'impartialité se retrouve nécessairement dans toute conception sérieuse de la justice. Le problème, c'est qu'une infinité de variantes de justice respectent ce critère, et certaines correspondront plus à l'intérêt privé de l'agent que d'autres (Elster 1999: 339). Cela ne signifie pas que l'agent choisit intentionnellement le concept de justice qui l'avantage le mieux (quoique la possibilité ne soit pas exclue), mais il est clair que la transmutation de l'intérêt en raison s'effectue plus aisément lorsque les deux exhibent des points communs. Avec le temps, une contrainte de cohérence s'impose : l'agent aura tendance à conserver les mêmes

---

[48] Quoiqu'un mécanisme secondaire, la "force civilisatrice de l'hypocrisie", peut faire dériver la fausse représentation vers une transmutation dans le cas des débats publics, en incorporant à la longue dans l'"image de soi" les déclarations faussement raisonnables de l'agent. Personnellement je crois que l'on surévalue grossièrement l'importance de ce mécanisme.

critères de justice dans différents contextes. On ne peut à la fois croire sincèrement en un principe quelconque de justice et le modifier ou l'abandonner selon qu'il sert nos intérêts privés ou pas, cela irait à l'encontre de notre estime de soi[49] (Elster 1999: 343-49). Le sentiment de honte associé à une baisse de notre estime de soi nous force en quelque sorte à respecter nos propres critères de justice.

La fausse représentation des préférences est également reliée aux normes sociales. Elle émane d'une pression sociale à se comporter correctement dans l'agora. Il est essentiel de noter que la norme conduit l'individu à *paraître* motivé par des considérations de justice, non à être sincèrement motivé. Il peut masquer ses préférences pour deux raisons principales, le conformisme et la persuasion[50] (ce qui correspond aux deux mécanismes délibératifs de formation de préférences impartiales, la censure de groupe et la persuasion). Cette première raison s'applique en vue d'éviter la désapprobation et le mépris d'autrui, causant respectivement de la culpabilité et de la honte, lorsque des propositions favorisant son propre intérêt sont présentées à une assemblée. Le simple fait de qualifier dès le départ une assemblée de "délibérative" - ou bien de marchandage, ou de simple vote - exerce une influence considérable sur le genre de propositions que l'on peut y émettre (Elster 1998b: 100). La persuasion ressemble beaucoup au conformisme; mais alors que ce dernier prend la forme négative d'une censure, la persuasion se sert de la raison comme avantage de négociation. Si l'agent croit qu'il sera plus en mesure d'obtenir ce qu'il veut de l'assemblée en maquillant ses propositions sous des traits raisonnables et impartiaux, alors il devient rationnel pour lui d'agir ainsi. Un cas intéressant concerne l'expression de menaces sous forme de mises en garde. Proférer une menace constitue une volonté d'usage de sa position de pouvoir, ce qui va à l'encontre de l'exigence de la "force du meilleur argument". En transformant intentionnellement

---

[49] Il ne s'agit pas ici de la force civilisatrice de l'hypocrisie, car la croyance n'est pas hypocrite.

[50] Si l'agent masque ses préférences par conformité, solidarité, etc., alors il intériorise les principes délibératifs que les autres attendent de lui et cela devient une forme de transmutation - en fait il ne les masque plus, il les transforme. Bien sûr, si la solidarité demeure hypocrite, le cas de tromperie demeure aussi. Nous pouvons toutefois imaginer toutes sortes de situations où il n'est pas clair si nous sommes en présence de l'un ou de l'autre de ces deux mécanismes, ou d'une combinaison perverse des deux, étant donné que tout dépend du profil psychologique intime de l'agent. Mais ce phénomène ne devrait pas nous empêcher d'élaborer des modèles de comportement valables, pas plus que les doutes soulevés quant au degré de rationalité des individus ne nous empêche de construire des modèles de choix rationnel valables. Il s'agit d'être vigilants.

une menace en mise en garde, celle-ci devient un argument raisonnable en bonne et due forme, pouvant être débattu (Elster 1998b: 100-4; Schelling 1960). Par exemple, au lieu de la menace du patron adressée à ses employés : "Si vous n'abandonnez pas vos revendications je vais effectuer des mises à pied", nous pourrions retrouver la mise en garde : "Si vous n'abandonnez pas vos revendications, le marché me forcera à effectuer des mises à pied". La première proposition tire sa validité de la crédibilité et de la position de pouvoir du locuteur, la seconde tire la sienne de la justesse de l'énoncé factuel.

En plus de la contrainte de cohérence, la fausse représentation fait face à une contrainte d'imperfection. La proposition raisonnable ne doit pas coïncider trop parfaitement avec les intérêts particuliers du locuteur, sinon on se doute de quelque chose. Idéalement l'agent devrait exprimer des positions raisonnables s'adressant à un large éventail de citoyens, pas seulement à lui et ses proches (Elster 1998b: 104; 1999: 376-77). Un autochtone réclamant des ressources pour lui et son peuple sera moins respecté qu'un autre réclamant ces ressources pour tous les autochtones; cela même si dans les deux cas, le locuteur obtient la même part pour lui-même. Ces deux contraintes favorisent la sincérité de l'argument. Évidemment, plus l'argument apparaît sincère, en fait, plus il aura l'air d'une transmutation réelle plutôt que d'une représentation hypocrite, mieux il sera accepté par l'assemblée.

Il existe une autre forme de fausse représentation, nullement motivée par les émotions et les normes sociales. Ce sont les préférences stratégiques. Ici, l'agent trouve avantage à ne pas exprimer sa préférence réelle afin d'infléchir le résultat collectif en sa faveur, en utilisant ainsi les particularités de la procédure, en se "jouant" d'elle. Considérons l'exemple suivant. On demande à trois individus (A, B, C) d'ordonner les options x, y et z en précisant que dans l'agrégation, le premier choix vaudra 4 points, le second 3 points et le dernier 1 point. Les ordres de préférences sont les suivants: A = [x, y, z], B = [x, y, z], C = [z, y, x]. Si l'individu C connaît les préférences des deux autres et qu'il est persuadé qu'ils voteront honnêtement, il sera alors rationnel pour lui d'exprimer l'ordre [y, z, x] car son vote "réel" donne comme résultat R_1 ($x = 9$, $y = 9$, $z = 6$) et le vote "stratégique" donne R_2 ($x = 9$, $y = 10$, $z = 5$). Comme C préfère réellement y à x et que z ne peut pas gagner, il devient rationnel de voter stratégiquement. Ce qu'il faut considérer, c'est que le choix de C ne se fonde pas uniquement sur l'ordre primaire

des alternatives, mais également sur l'ordre secondaire des résultats possibles anticipés, soit $[R_2, R_1]$.

Le fardeau de la responsabilité représente une troisième forme de fausse représentation : je peux préférer une politique particulièrement pénible à appliquer sans vouloir passer pour celui par qui le malheur arrive. Dans un même ordre d'idée, les méthodes de votes à découvert (ou "main levée") peuvent inciter à l'hypocrisie; c'est pour cette raison que les votes secrets sont souvent privilégiés. Ce genre de comportement hypocrite survient en général lorsque le geste de voter ou de s'exprimer comporte des conséquences au-delà du débat social en cours. Pour revenir au vote à main levée, ce geste, en plus de signaler une prise de position, a une influence certaine sur les relations futures avec les concitoyens pouvant observer le geste, conséquences qui se répercuteront en d'autres circonstances. Par conséquent, l'électeur rationnel votant de cette façon doit considérer dans son calcul d'utilité non seulement le résultat du processus de décision, mais aussi les externalités découlant du simple geste, ce qui peut faire pencher la balance en faveur d'une fausse représentation de son choix.

La démocratie délibérative n'accepte pas la fausse représentation, en particulier le vote stratégique. Elle y voit une insincérité et un manquement général aux préceptes du citoyen raisonnable et accuse fréquemment la théorie du choix social de permettre ce genre de comportement. En revanche, plusieurs théoriciens du choix social se sentent mal à l'aise avec ce phénomène et cherchent à en minimiser la portée. Il faut d'abord se demander en quoi le vote stratégique discrédite le choix social. La théorie se fonde sur une conception rationaliste de l'individu et le vote stratégique correspond parfaitement au type de comportement auquel on doit s'attendre d'un tel individu. Dans la plupart des cas, l'apparition de ce phénomène nous rappelle simplement que le geste de voter ne se déroule pas en vase clos. Pour conserver sa cohérence, la théorie devrait accepter toutes les conséquences découlant du choix rationnel, incluant les ordres secondaires de préférences et les externalités. La démocratie délibérative n'a pas non plus à rejeter la fausse représentation. Des trois types mentionnés, le premier peut inciter, de façon certes imparfaite, le comportement raisonnable par des normes sociales

d'impartialité et de civisme; le second peut contribuer à contrer des alternatives inacceptables[51] et le troisième est inévitable pour les théories qui s'intéressent au fonctionnement de la délibération publique et transparente.

Nous pourrions rapprocher le concept de normes sociales de celui de la "rationalité de rôle" (Goodin 1986: 88-89) stipulant que nos critères d'évaluation rationnelle dépendent du contexte; dans le cas présent, il y aurait une distinction entre une rationalité "politique" exercée dans l'agora et une rationalité plus instrumentale exercée dans le marché. Il s'agit d'une forme d'extension du choix rationnel orthodoxe, spécifiant que la satisfaction de préférences égoïstes n'est pas la seule avenue possible et que les agents sont suffisamment intelligents pour réfléchir en termes bénéfiques pour la société lorsqu'ils se retrouvent dans une situation décisionnelle pouvant affecter cette société. Plusieurs théoriciens se sont penchés sur cette rationalité politique. Pour Sunstein (1993: 208-9), les "préférences politiques" respectent les aspirations collectives et ont un penchant pour l'altruisme et la sympathie. Elles peuvent également prendre la forme de "méta-préférences", des préférences pour des types de préférences (vouloir être impartial, par exemple). Boudon (1998: 191) offre le concept de "rationalité axiologique" : sous certaines circonstances, la raison se fonde sur des critères non conséquentialistes de moralité et de justice. On retrouve chez Harsanyi (1990: 278) une théorie générale tripartite du comportement rationnel, se subdivisant en théorie de l'utilité, théorie des jeux et théorie éthique. Cette dernière se veut une théorie des jugements de valeurs moraux rationnels. Les préférences y sont fondées sur l'impartialité et l'agent cherche à maximiser l'utilité moyenne de la collectivité. Enfin, Sen cherche à compléter la théorie du choix rationnel par les concepts de "sympathie", l'inclusion de l'utilité d'autrui dans sa propre utilité, et d'"engagement", soit les considérations morales dans la formation de l'utilité. Pour lui l'usage de la moralité demeure contextuel, différentes problématiques collectives amèneront différentes positions morales (Sen 1982: 91-93, 98-99).

[51] Prenons exemple sur les élections présidentielles françaises de 2002. Au premier tour, un électeur de gauche qui n'avait pas l'intention de voter pour Lionel Jospin peut décider de "faussement" voter pour lui lorsqu'il apprend que Jean-Marie Le Pen a des chances de passer au second tour. Ce geste a toutes les allures d'un vote stratégique et pourtant, il est moralement irréprochable.

Un autre phénomène favorisant la discussion raisonnable consiste en ce qu'on pourrait nommer le "voile d'ignorance temporel" : même si les individus connaissent leur position sociale, leurs ressources, etc., une incertitude demeure toujours quant à l'avenir, ce qui les force jusqu'à un certain point à adopter des positions ne favorisant pas leur statut au détriment des autres. Ce point est encore plus vrai lorsque l'individu prend à coeur les intérêts de ses proches et de sa descendance. Nous obtenons une coïncidence entre l'intérêt particulier (possiblement altruiste) et l'impartialité (Shepsle 1989: 138-39; Elster 1998b: 114-16).

Les comportements stratégiques

Le qualificatif "stratégique" désigne les actions individuelles, entreprises dans le cadre d'une décision collective, autres que le geste direct de voter ou de donner son opinion comme tel. Ces actions prennent une tournure indirecte, souvent considérée comme insidieuse. Il n'en est pas toujours ainsi. Avec le phénomène de la fausse représentation, nous avons vu que les comportements stratégiques ne sont pas nécessairement incompatibles avec la recherche collective de la bonne société. En plus du vote stratégique consistant en une réorganisation par l'individu de ses préférences sur la base des résultats agrégés anticipés, nous allons maintenant nous intéresser à deux autres formes de stratégie : le choix de la procédure et le marchandage de votes.

Le phénomène d'*ambiguïté*, bien connu dans la littérature du choix social, stipule qu'un même ensemble de choix individuels peut générer des résultats collectifs différents tout dépendant de la procédure adoptée, ce qui créé des opportunités de manipulation de la part de ceux contrôlant l'ordre du jour politique. Même des citoyens parfaitement raisonnables se retrouveront aux prises avec cette difficulté car il n'existe pas de procédure parfaitement objective, mais dans leur cas on ne suppose pas que les intérêts personnels auront une influence sur le choix. Nous rencontrons dans les régimes démocratiques des procédures majoritaires dont le seuil de validité varie en fonction de l'importance de l'enjeu; par exemple, la majorité requise pour un amendement constitutionnel est plus élevée que pour l'adoption de lois ordinaires. On peut y voir un certaine sensibilité morale à vouloir conférer une importance spéciale aux institutions fondatrices libérales. Une autre explication, compatible avec la précédente, serait de protéger stratégiquement, selon le principe rationnel du "voile d'ignorance

temporel", ces institutions contre l'exploitation par les générations suivantes (Elster 2000, ch. II). On s'aperçoit que le choix stratégique n'est pas nécessairement immoral.

L'ambiguïté causée par la multiplicité des procédures agrégatives peut se voir en partie atténuée par une délibération portant sur la procédure particulière de vote permettant de trancher les débats si, bien sûr, les participants décident de passer au vote plutôt que de rechercher le consensus à tout prix. Ceci n'élimine pas le problème fondamental de l'ambiguïté mais permet au moins une justification reconnue par tous de la procédure adoptée. La justification demande toutefois que les individus placent le respect de l'idéal démocratique au-dessus de leurs intérêts personnels (Miller 1992: 66).

Un cas intéressant concerne la distinction entre procédures agrégatives portant sur des prémisses et sur une conclusion (Pettit 2001). Un exemple : un syndicat doit se décider sur le déclenchement d'une grève en spécifiant que si les membres jugent à la fois que les salaires sont trop bas et que les conditions de travail sont inadéquates (les prémisses), alors le mandat de grève sera valide (la conclusion). Supposons que 40% des membres croient que les salaires sont trop bas mais que les conditions sont adéquates, qu'un autre 40% croient exactement le contraire, et que les 20% restant croient les deux prémisses vraies. Pettit démontre aisément que si le vote porte sur les prémisses, soit deux votes distincts et une conclusion dérivée des résultats, le résultat favorisera la grève (car chaque prémisse remportera 60%) tandis que si le vote porte sur la conclusion, la grève ne passera pas, remportant seulement 20%. Il est clair que le choix de la procédure est ici crucial. Laquelle est la plus juste ? D'après Pettit, le vote sur prémisses semble plus juste car il respecte un critère fondamental (selon lui) de la démocratie délibérative, le débat sur des positions communes connues de tous[52] (Pettit 2001: 2). Mais on peut le contrer en stipulant que seul 20% des membres désirent vraiment la grève. L'utilité d'un vote sur prémisse apparaît lorsque le groupe a besoin, en plus de répondre à une question affectant ses membres, de se constituer une "raison collective" ou, en d'autres termes, une ligne de parti. Le groupe peut ainsi offrir au public les raisons de sa décision. Les mérites

[52] Dans le même ordre d'idée, le vote sur conclusion incite à la "paresse" car il n'exige pas que l'individu fournisse les raisons de son choix (Pettit 2001: 22). Cette critique porte sur les prémisses disjonctives, lorsqu'une seule prémisse vraie est suffisante pour que la conclusion soit vraie.

de chacune des procédures dépendra du contexte et des buts visés par le groupe, et bien entendu la possibilité de manipulation stratégique demeure. Pettit ne réussit pas à nous convaincre que le vote sur prémisses est *a priori* préférable au vote sur conclusion, toutefois sa distinction demeure pertinente car on rencontre fréquemment ce genre de choix en politique.

L'échange de vote représente un autre champ stratégique en démocratie. Il n'est pas nécessaire que les votes changent littéralement de mains, il s'agit simplement que les électeurs penchent d'un côté ou de l'autre selon certains incitatifs reçus. L'opinion populaire condamne ce genre de transaction; elle violerait le principe "une personne, un vote" et elle ouvrirait la porte aux abus de pouvoir, entre autres. Bien qu'en général cette critique soit attrayante, il existe un argument soulignant le caractère potentiellement équitable d'un tel échange. Dans sa conception "distributive" de la procédure, Christiano (1993: 182-84) propose que la recherche de l'égalité ne s'effectue pas au niveau de la procédure de vote, mais bien au niveau du *processus* global, soit le vote avec tout ce qui l'entoure (collecte d'information, coalitions, etc.). Tout comme l'égalité économique ne signifie pas que chacun possède exactement la même quantité de chaque bien mais plutôt une égalité globale sur l'ensemble des biens, l'égalité politique ne devrait pas porter sur l'égalité à chaque application de la procédure de vote mais plutôt sur l'égalité globale de l'ensemble des décisions collectives. Il serait donc plus juste de permettre aux citoyens de marchander leur vote sur des sujets qui ne les intéressent pas. Toutefois, la justesse de ce principe dépend d'une distribution égalitaire des "ressources politiques", au dire même de l'auteur, afin d'éviter la possibilité d'abus de pouvoir. Sans la présence d'institutions prévenant ces abus et régulant le tout, cette forme d'égalité se rapproche de la "main invisible" en économie. Toutefois, l'échange de votes demeure un fait politique commun, même parmi des citoyens raisonnables, et ce phénomène n'est pas nécessairement injuste.

La délibération publique peut aussi être utilisée stratégiquement. Bien sûr, dans un tel cas elle ne pourra être qualifiée de "raisonnable", l'usage stratégique de la délibération allant à l'encontre de tout ce que la démocratie délibérative propose. L'adoption de principes en public peut contribuer à une sympathie envers autrui et une volonté de remettre en question ses principes au nom de la recherche du consensus, c'est ce que la démocratie délibérative suggère,

mais elle peut aussi provoquer exactement le contraire, soit un durcissement des positions. Le locuteur y met en jeu son honneur et sa réputation, c'est ce qui permet un tel durcissement (Elster 1994a). Ce phénomène se veut particulièrement utile lors de situations présentant plusieurs forums, par exemple une délibération comprenant des moments à huis clos et en public. L'individu peut grandement améliorer sa position de négociation à huis clos en s'engageant publiquement au préalable à respecter certains principes. Dans une perspective de justice, l'engagement public peut constituer un instrument précieux de défense entre les mains de ceux occupant une position défavorable dans un débat inégal : "Quand les puissants négocient en s'appuyant sur la force, les faibles le font en s'appuyant sur des principes" (Elster 1994a: 244).

Cette section se voulait un bref survol des mécanismes pouvant engendrer la délibération en contexte rationnel. Nous avons voulu montrer que la formation et l'expression de préférences minimalement délibératives était possible chez des citoyens rationnels avec des mécanismes tels la transmutation, la fausse représentation, la rationalité de rôle ou encore le voile d'ignorance temporel. Nous avons remarqué aussi que les préférences stratégiques, un comportement rationnel inévitable, pouvaient avoir des conséquences positives pour la délibération raisonnable. Ces mécanismes sociaux nous ont permis d'entrevoir un vaste domaine sous-estimé autant par les théoriciens de la démocratie délibérative que du choix social. Il y a encore beaucoup à dire sur la rationalité politique, les stratégies argumentatives, les relations de pouvoir, et bien plus encore. Tout de même, ce survol permet de nous rendre compte de l'importance capitale de ce champ d'étude de la démocratie que sont les interactions sociales en contexte mixte de rationalité, de normes sociales et de moralité.

2.3 Conclusion

Nous avons constaté que la démocratie délibérative avait peine à appliquer les principes de justice sociale qu'elle propose. Élaborer des thèses de justice, aussi correctes soient-elles, n'est pas la même chose que d'en penser la mise en application, même dans des modèles abstraits. La démocratie délibérative a tendance à évacuer un peu trop rapidement les problèmes d'implémentation pour se concentrer sur les discussions de justice, ce qui constitue une lacune pour une théorie cherchant à créer une nouvelle façon de gouverner.

La discussion sur la délibération circonstancielle nous a permis d'établir la possibilité de délibération parmi des individus rationnels. Alors que de nombreux démocrates délibératifs relèguent les questions de formation et d'expression de préférences "raisonnables" à un mécanisme simpliste des vertus éducatrices et civilisatrices de la délibération, le modèle motivationnel de Elster va beaucoup plus loin et nous permet de découvrir un univers complexe de normes personnelles et sociales, et de comportements plus ou moins sincères. D'autres principes comme la rationalité de rôle et le "voile d'ignorance temporel" viennent préciser cette possibilité de délibération. De plus, certains comportements rationnels n'ayant que peu de familiarité *a priori* avec la délibération comme tel n'en demeurent pas moins compatibles avec les préceptes de la démocratie délibérative. C'est le cas notamment du vote stratégique, de la fausse représentation induite par le fardeau de la responsabilité et du choix de la procédure de vote lorsqu'un tel choix devient inévitable. La leçon à retenir ici est que le geste de délibérer (ou de voter) se déroule dans un contexte particulier qu'on ne peut ignorer. Nous avons vu avec l'exemple des présidentielles françaises que le vote stratégique est parfois nécessaire à la préservation d'une société juste. Le fardeau de la responsabilité nous invite à considérer les externalités liées à l'acte de délibération. La procédure "juste" de délibération ou d'agrégation varie d'une situation à l'autre.

Nous n'avons pas voulu ici placer les principes de la démocratie délibérative au banc des accusés. Il ne s'agit nullement d'une critique au niveau des principes de justice, mais si la démocratie délibérative désire appliquer ses principes, alors elle doit prendre au sérieux les mécanismes sociaux rationnels que nous avons proposés. On ne peut admettre du même souffle des effets pervers de groupe (qui ne relèvent pas simplement de l'égoïsme) et souhaiter qu'ils disparaissent avec des citoyens raisonnables. Rawls semble l'avoir compris jusqu'à un certain point, c'est pour cette raison qu'il sépare son modèle idéal de "position originelle" de la discussion sur l'application de ce modèle, comprenant de nombreux postulats socio-politiques. Plusieurs démocrates délibératifs (Cooke 2000: 957-67; Benhabib 1996: 75) lui ont reproché de n'appliquer son modèle que dans les cas constitutionnels essentiels. Il faut peut-être y voir une certaine prudence théorique. En conclusion, la délibération circonstancielle fait apparaître la théorie large de la rationalité comme une théorie de l'individu à la fois utile pour la mise en

application des principes de justice et pas forcément anti-délibérative. La distinction entre l'individu rationnel et le citoyen raisonnable n'y apparaît plus aussi claire.

CHAPITRE III

LES FONDEMENTS RATIONNEL ET ÉMOTIF DES NORMES SOCIALES

Le phénomène bien connu de l'agent supposé rationnel se conformant à des normes sociales représente un casse-tête de longue date pour la théorie du choix rationnel. Nous pouvons dégager deux grandes conceptions des normes présentes dans cette théorie. D'abord, l'agent peut se conformer à la norme de façon tout à fait rationnelle, soit dans le but d'éviter les sanctions sociales ou pour réclamer sa part du bien commun produit par le respect général de la norme. Mais il y a des normes qui ne produisent aucun bien commun. Aussi, la logique des sanctions nous entraîne vers une régression infinie (qui sanctionne le sanctionneur ?). Plus encore, on observe fréquemment des individus se conformant aux normes sans qu'il y ait possibilité de sanctions. La seconde stratégie est alors de fonder les normes sur une motivation de nature émotionnelle : leur violation entraîne un sentiment désagréable de honte ou de culpabilité chez l'agent. Il s'agit ensuite de modéliser ce désagrément comme une "désutilité", qui vient s'ajouter naturellement à la fonction d'utilité. Une variante similaire consiste à modéliser les normes comme préférences et à concevoir un ordre de "méta-préférences" morales réglant l'ordonnancement des préférences "ordinaires". Là encore, il y a problème. Les émotions provoquées par les normes ne sont pas seulement sources d'inconfort, elles peuvent également perturber et même bloquer la faculté de faire des choix rationnels. La solution proposée par Elster, dont nous avons pris connaissance au chapitre I, consiste à distinguer les motivations rationnelles, fondées sur la satisfaction de ses intérêts, et les motivations non rationnelles, fondées sur les émotions. Elster propose que le champ d'étude de cette dernière catégorie se situe à l'extérieur de la théorie du choix rationnel, et qu'elle s'appuie plutôt sur la psychologie et la sociologie. Le présent chapitre se veut une discussion accompagnée d'une tentative de reconstruction de son modèle. L'objectif est de pouvoir formuler des motivations de nature émotionnelle à se conformer aux normes sociales, sans pour autant écarter les motivations rationnelles.

Nous établissons d'emblée une distinction entre normes personnelles et normes sociales. Une norme personnelle est une règle de conduite spécifique à l'agent, dont la conformité est soutenue par les émotions de honte et de culpabilité. Une norme sociale est une règle de conduite externe, partagée par les membres de la communauté, habituellement soutenue par un système de sanctions, dont la caractéristique fondamentale est qu'elle exige de chacun qu'il donne au minimum l'apparence qu'il sacrifie une partie de son intérêt particulier au nom de l'intérêt de la communauté[53]. Dans le cadre de la relation entre rationalité et normes, nous proposons une qualification de cette caractéristique : l'agent ne doit pas paraître respecter la norme par pur calcul rationnel. Cette exigence de sacrifice exclut ce qu'on appelle communément les conventions, soit les règles servant à résoudre directement des problèmes d'action collective, comme rouler à droite sur les routes, par exemple. En effet, dans le cas des conventions, l'action normativement correcte se confond avec la maximisation de l'intérêt. Dans notre modèle, une norme sociale est intériorisée lorsqu'elle correspond à une norme personnelle. L'agent se trouve motivé à se conformer à une norme sociale lorsque les valeurs exprimées par celle-ci correspondent aux valeurs de l'agent exprimées cette fois-ci par ses normes personnelles. L'importance d'une norme personnelle pour l'individu influencera son désir de se conformer à la norme sociale, et déterminera également laquelle des émotions, honte ou culpabilité, sera déclenchée en cas de violation.

Nous postulons également trois désirs fondamentaux, présents à divers degrés chez tout individu : la promotion de son intérêt personnel, le maintien d'une image positive de soi et l'inclusion sociale. Ce dernier sera perçu soit comme une norme personnelle lorsque la solidarité représente une valeur pour l'agent, soit comme une préférence lorsque celle-ci est pour lui une source d'utilité. La solidarité comme préférence implique que l'agent puisse se conformer rationnellement à une norme sociale lorsque l'utilité intrinsèque des bonnes relations avec ses pairs l'emporte sur les coûts du conformisme.

Nous ne visons pas ici à une contribution à la théorie psychologique des émotions. Il se veut plutôt une discussion sur la place de la honte et de la culpabilité comme motivations à

---

[53] Cette caractéristique est fortement inspirée de Pierre Bourdieu. Elle figure dans plusieurs de ses écrits, la discussion la plus explicite se trouvant dans l'essai "Un fondement paradoxal de la morale" (Bourdieu 1994: 233-238).

l'action dans le cadre de la théorie du choix rationnel, en particulier dans leur relation avec les normes sociales. Notre but est d'en arriver à un modèle simple du fondement des normes sociales intégrant rationalité et émotions. Pour ce faire, nous devrons adapter les concepts de honte et de culpabilité à notre problématique spécifique. Il est entendu que ces concepts demeurent beaucoup plus riches que l'utilisation que nous en faisons. Les émotions de culpabilité et de honte seront étudiées ici avant tout comme sentiments négatifs déclenchés par la violation de normes. Nous débuterons par une définition primaire, qui correspondra à une certaine définition générale retenue par la psychologie contemporaine et qui se marie bien avec notre propos. Par la suite, nous passerons à une définition opérationnelle, où nous tenterons d'intégrer les concepts de la définition primaire à notre théorie de la rationalité. Dans la troisième partie, nous orienterons nos efforts vers l'anticipation de l'émotion et les réactions possibles de l'agent cherchant à éviter ces émotions négatives. Le cas le plus intéressant survient lorsque l'agent se conforme aux normes en redécrivant ses intentions plutôt qu'en renonçant à ses buts; ce sera le sujet de la dernière partie.

3.1 Définition primaire

La honte et la culpabilité font partie de la classe des émotions évaluatives, qui impliquent une évaluation de l'objet de l'émotion (Elster 1999: 143; M. Lewis 1993: 566-67). Cette classe se subdivise en trois catégories : l'émotion peut porter sur soi ou sur autrui, son objet peut être la personne même (évaluation globale) ou une action (évaluation spécifique), et elle peut être positive ou négative. La honte et la culpabilité sont toutes deux des émotions négatives portant sur soi. Elles se distinguent par leur objet. La définition la plus simple sera donc la suivante : la honte est le résultat d'une auto-évaluation négative de sa personne alors que la culpabilité est le résultat d'une auto-évaluation négative d'un geste commis. On retrouve l'origine de la distinction entre les deux émotions sur la base de leur objet chez H. Lewis (1971). Bien que son ouvrage soit de nature psychanalytique, sa distinction originale a été bien reçue par de nombreuses écoles psychologiques et a depuis largement influencé les travaux sur ces émotions[54] (Tangney et al. 1996: 1257).

[54] En effet, la grande majorité des textes utilisés ici s'y réfère explicitement. Tangney a démontré à de nombreuses reprises la validité empirique de cette distinction.

La honte est donc une atteinte directe à l'estime de soi. C'est ce que l'agent ressent lorsqu'il viole une norme qui lui est fondamentale : "(...) it is less shame about transgressing some rule or violating some sanction but rather about failing to meet or approximate one's own moral ideal" (Manion 2002: 77). La honte peut être déclenchée de deux manières, par la violation d'une norme personnelle immédiatement perçue comme telle par le sujet, et par le mépris d'autrui; nous parlerons respectivement de honte *personnelle* et de honte *sociale*. Cette dernière est d'ordinaire plus intensément ressentie que la honte personnelle, car notre image publique se voit rabaissée en plus de notre estime de soi. Alors que la honte personnelle peut se vivre comme un "fantasme" (*fantasy*) de dégradation (Jacoby 1994: 3), dans la honte sociale, la dégradation est ressentie de manière bien réelle. Rawls (1971: 444) utilise les termes "honte naturelle" et "honte morale" pour désigner respectivement ces deux types. Pour lui, tous deux résultent d'un échec moral personnel mais la honte morale résulte en plus d'un échec aux yeux des autres. On retrouve chez Klaassen (2001: 180) une distinction semblable : la honte est une dégradation de l'image de soi alors que l'humiliation est une dégradation de son image publique. Ceci demeure toutefois compatible avec notre modèle, car il précise qu'une humiliation "convaincante" va déclencher de la honte. Donc, la honte est en définitive personnelle, et la honte sociale ne fait que rajouter un déterminant supplémentaire. Dans une perspective fonctionnaliste, la honte (ainsi que la culpabilité) sert à satisfaire le besoin d'inclusion sociale en forçant le respect de normes favorisant la vie en société (Tangney et al. 1996: 1267-8). Jacoby adopte également cette fonction sociale mais en rajoute une autre, la protection de la sphère intime de l'individu. Dans cette perspective, la honte permet de préserver sa "distinction", son identité propre, créant cette gêne qui empêche de se révéler tout entier au monde. Toutefois, Jacoby précise que même la distinction a une origine sociale : "(...) it is society that demands a particular degree of discretion from each individual (...). Society decides what befits the individual, what is proper according to social mores." (Jacoby 1994: 21).

Alors que, dans ces théories, la honte porte sur la personne, la culpabilité porte sur les actions. Alternativement, selon la formule de Barrett (1995: 43), la honte constitue une évaluation de soi comme objet et la culpabilité une évaluation de soi comme sujet. La culpabilité survient lorsque l'agent commet un acte à l'encontre de ses propres valeurs et qui a

pour effet de porter atteinte à autrui, et pour lequel il se sent responsable[55] (Rawls 1971: 445-46, 484; Manion 2002: 76). Pour Ogien (2002: 30-31), c'est cette admission de responsabilité qui distingue la culpabilité; dans le cas de la honte, une telle admission demeure ambiguë et en quelque sorte, refoulée. Enfin, pour Gibbard (1990: 139-40), la culpabilité se veut une réponse à la colère et aux sanctions d'autrui, alors que la honte constitue une réaction au rejet et au mépris. La fonction sociale de la culpabilité consisterait en l'amélioration des relations interpersonnelles à travers l'activité réparatrice (Tangney et al. 1996: 1267). Dans la définition rawlsienne, il n'est pas nécessaire que la victime réagisse contre l'agent fautif pour déclencher de la culpabilité : celui-ci peut demeurer incognito tout en se sentant coupable d'avoir mal agi. Tout comme dans le cas de la honte, nous retrouvons une variante personnelle et une variante sociale de l'émotion, cette dernière se manifestant généralement plus intensément.

Comme ce qui nous intéresse, c'est l'*action* individuelle et l'influence possible des émotions, la distinction évaluation de soi/évaluation de l'action prendra une autre forme. Dans un contexte d'agent, la distinction ne se pose plus aussi clairement. Dans un sens, toute émotion portant sur une action doit en définitive porter sur "soi". La critique d'autrui d'une action commise qui n'atteint pas le "soi" de l'agent ne déclenchera aucune émotion. C'est justement cette atteinte de soi à travers l'action qui met en jeu les émotions. L'évaluation négative d'une action se résout dans l'évaluation négative de soi, car l'admission d'une responsabilité de la transgression révèle déjà un défaut de caractère (Sabini, Silver 1997: 5). Si nous postulons que les deux émotions partagent la même origine, l'atteinte à l'estime de soi, qu'est-ce qui les distingue alors ? Il est concevable que nous ayons affaire à un phénomène continu, où la honte et la culpabilité se situeraient aux pôles et où la frontière entre les deux prendrait l'allure d'une zone grise. Bien qu'ils aient tenté de les distinguer, la plupart des auteurs cités plus haut reconnaissent qu'un même phénomène peut générer de la honte ou de la culpabilité, et qu'il existe très peu de situations canoniques. Au niveau présent des connaissances, nous en sommes réduit à établir des distinctions contextuelles qui ne peuvent être universalisées[56]. Pour notre modèle, nous allons adopter la thèse du continuum, dont le

[55] L'admission publique de responsabilité est une toute autre paire de manches, comme nous le verrons plus loin.

[56] Voir les critiques de Gibbard (1990) et de Ogien (2002) des différentes tentatives de distinction dans la littérature.

vecteur représente l'intensité des normes personnelles en jeu et l'élément déclencheur, une auto-évaluation de soi, possiblement provoquée de l'extérieur. Sans y adhérer totalement, Rawls penche dans cette direction lorsqu'il relie la culpabilité aux valeurs de justice et la honte aux "excellences" de l'individu et à son autonomie (*self-command*). Également, la plupart des auteurs s'entendent sur le fait que la culpabilité est une émotion moins intense que la honte.

Dans l'autre sens, étant donné notre intérêt postulé pour l'action, nous excluons pour fins de modélisation les atteintes au "soi" qui n'originent pas d'une action responsable quelconque. En outre, nous ne traiterons pas d'un certain aspect de la honte, à savoir, celui concernant les états pour lesquels nous n'avons aucun contrôle, comme avoir honte de sa laideur par exemple. Bien que le débat reste ouvert entre théoriciens de la psychologie, nous considérerons cette forme de honte "irrationnelle" ou "inappropriée"[57] (Manion 2002: 73-74). Sans vouloir nous attarder sur le sujet, il est à noter que cette honte peut devenir appropriée si elle prend une forme sociale, par exemple, lorsque quelqu'un exprime du mépris pour la laideur d'un autre (Elster 1999: 150). Dans la définition qui nous intéresse, l'agent doit conserver un certain contrôle sur ses actions pouvant déclencher la culpabilité ou la honte; en d'autres termes, on doit pouvoir le tenir pour le moins responsable. L'atteinte au "soi" consiste en fait en une atteinte à une norme personnelle. A dit à B: "Ce que tu as fait est mal!". Pour que B ressente quelque chose, il faut que l'action viole en effet une norme personnelle[58]. La simple accusation n'est pas suffisante. Nous exigeons donc dans notre modèle un ensemble de relations entre l'évaluation négative de ses actions et les normes personnelles. L'accusation plus globale: "Tu es une mauvaise personne!", censée être le *locus* de la honte, doit pour nous être engendrée par une action: "Tu es une mauvaise personne *parce que* tu as fait ceci...". La forme de l'accusation influencera bien sûr le degré de l'émotion ressentie par la victime. En fait, même à l'extérieur de notre modèle restrictif, il est difficile d'imaginer une accusation globale qui ne

---

[57] Les termes "rationnel" et "irrationnel" concernant les émotions apparaissent fréquemment dans la littérature psychologique. Puisqu'il est ici largement question de choix rationnel, nous éviterons la confusion en y substituant les termes "approprié" et "inapproprié". En fait, c'est ce que les psychologues veulent dire : une émotion irrationnelle est une émotion inappropriée par rapport à la situation, qui n'a pas de raisons de survenir.

[58] La norme personnelle peut simplement être de ne pas déplaire à autrui, cette norme étant plus ou moins forte selon la "distance sociale" entre les sujets. Ainsi, l'accusation de la part d'un proche peut générer de la honte même si l'agent ne sait pas quelle norme sociale son action a violée. Je remercie Paul Dumouchel de m'avoir éclairé sur ce point.

soit pas reliée d'une façon ou d'une autre à l'action. Une exception importante serait le préjugé[59] Nous posons donc comme postulat méthodologique (i) que l'auto-évaluation de nos actes s'effectue à l'aune des nos normes personnelles et (ii) qu'il existe un lien causal entre la violation d'une telle norme et l'émotion, soit de honte ou de culpabilité.

Dans notre modèle, la honte et la culpabilité partagent les mêmes causes, la violation d'une norme personnelle, mais c'est le degré d'intériorisation de la norme (soit sa correspondance avec des valeurs subjectivement plus ou moins fondamentales) qui détermine le déclenchement de l'une ou l'autre. Celles-ci exhibent des tendances à l'action distinctes. Celle concernant la honte la plus communément relevée par les psychologues est un complexe de réactions immédiates incluant la fuite, le retrait, la paralysie, l'agressivité, etc. Sans nier la force souvent irrésistible de ce type de réaction, deux autres réactions positives sont tout de même possibles face à la honte, particulièrement après que l'intense malaise immédiat se soit estompé. Il y a d'abord la révision de ses idéaux moraux (Manion 2002: 84). Ici l'agent n'accepte pas le message véhiculé par le sentiment de honte et entreprend de se distancier de certaines valeurs de façon à ce qu'une situation similaire dans le futur ne déclenche pas à nouveau de la honte. On peut imaginer par exemple que les homosexuels qui "sortent du placard" passent par un tel type de révision. Il faut remarquer que pour toutes sortes de raisons trop élaborées pour être présentées dans cette étude, aucun agent ne peut transformer ses valeurs morales à volonté, sous peine de perdre le sens de son identité[60]. Il existe une force inertielle déterminante qui rend très difficile - mais pas impossible - une telle révolution. Lorsque l'agent accepte le jugement sur sa personne engendré par la honte, nous avons la seconde réaction non pathologique possible, le conformisme. C'est la méthode la plus efficace de se débarrasser de cette douloureuse émotion (Klaassen 2001: 186-7). La honte induit une "connaissance de sa place" en société et force l'agent à y demeurer (Greenspan 1993: 56). Dans la perspective des motivations possibles face aux normes sociales, nous proposons le conformisme comme principale tendance à l'action de la honte. Nous reviendrons plus longuement sur le sujet lorsque nous discuterons de l'anticipation de la honte.

---

[59] Même le préjugé est fréquemment redécrit en termes d'actions, vraisemblables ou non : "Ils débarquent dans notre pays pour voler nos emplois".

[60] A ce sujet, voir Livet (2002).

Les tendances à l'action immédiates de la culpabilité sont de nature plutôt "proactive", contrairement à la honte qui favorise le retrait et l'inaction. Ici, outre le conformisme, l'agent cherche à obtenir de nouveau la faveur d'autrui par l'aveu et l'activité réparatrice. L'agent peut également tenter de combattre ce sentiment. Suivant Miceli et Castelfranchi (1998: 311), nous dirons de l'agent niant sa responsabilité de l'acte fautif qu'il cherche des *excuses*, et de celui contestant l'évaluation négative de l'acte qu'il se *justifie*. Le jeu des excuses et des justifications est permis sous l'emprise de la culpabilité, mais pas avec la honte, car ici l'emprise de l'émotion est suffisamment faible pour permettre à l'agent de tenter de la contourner rationnellement. L'émotion de la honte est tellement prenante qu'elle bloque la rationalité, du moins immédiatement. C'est là une distinction cruciale pour notre modèle.

En conclusion, notre définition primaire pourrait se résumer ainsi. La culpabilité et la honte résultent d'une auto-évaluation négative de soi se fondant sur les normes personnelles. Nous postulons une relation nécessaire entre une telle violation et l'expérience de l'émotion, car c'est précisément ce qui maintient la norme; on se demande quelle serait la force d'une norme personnelle que l'on pourrait violer sans conséquences psychiques négatives. Par souci de méthode, nous avons choisi de nous concentrer sur l'action individuelle; nous établissons par conséquent une seconde relation, cette fois-ci entre l'évaluation négative par autrui de l'action et la violation de normes personnelles. Cette relation n'est toutefois pas nécessaire. Il est possible qu'une critique de l'action n'atteigne pas l'agent au niveau de ses normes. Une atteinte à l'estime de soi non provoquée par un acte quelconque est également possible, mais nous avons choisi ici de ne pas nous en préoccuper. Bien que culpabilité et honte soient deux émotions exprimant à des degrés variables le même phénomène psychique, une atteinte à l'estime de soi, il n'en demeure pas moins qu'elles se distinguent nettement dans leur tendance à l'action. Dans ce modèle, l'action individuelle fautive entraîne (possiblement) la violation d'une norme personnelle, déclenchant une des deux émotions, qui à son tour provoquera l'agent à réagir d'une manière propre à l'émotion. La faute peut être attribuée par l'agent lui-même ou par autrui, cette dernière cause se révélant généralement plus efficace.

Maintenant, comment intégrer les normes sociales dans ce schéma ? Ici, les normes sociales prennent l'aspect exogène de la règle, et l'aspect endogène des normes personnelles. L'intériorisation des normes sociales survient lorsque les valeurs sous-tendant la norme sociale correspondent aux valeurs propres à l'agent, ou lorsque le désir fondamental d'inclusion sociale pousse l'agent à adopter ces valeurs pour lui-même. Ce troisième type de relation, des normes sociales vers les normes personnelles, nous permet de déterminer, *pour chaque individu*, si la violation de la règle entraînera de la culpabilité, de la honte, autre chose, ou rien du tout. La primauté des normes personnelles sur les normes sociales se confirme par le fait que certaines normes personnelles sont telles qu'une action entreprise en leur nom et subissant l'opprobre d'autrui fondé sur des normes sociales ne générera pas de honte ou de culpabilité, mais de la colère et de l'indignation. Ces mêmes émotions peuvent également survenir lorsque l'agent transgresse une norme sociale qu'il considère inutile ou dépassée et que quelqu'un tente de le remettre à sa place. Les normes sociales provenant de l'extérieur, nous avons la capacité, même instinctive, de les évaluer, et le résultat de cette évaluation déterminera le genre d'émotions qu'éveillera leur violation. Il n'en est pas de même pour les normes personnelles, qui sont complètement intériorisées; leur violation entraînera un certain degré de honte ou de culpabilité selon l'importance qu'elles présentent.

3.2 Définition opérationnelle

Armés de cette conceptualisation primaire, nous allons aborder l'intégration de ces deux émotions dans un modèle de choix rationnel qui nous mènera par la suite à une théorie des normes sociales. Dans son ouvrage *Alchemies of the Mind* (1999), Elster nous offre une telle définition opérationnelle qui apparaît, selon moi, adéquate mais présentant certaines imperfections. Nous procéderons par l'étude de sa définition et, à partir des critiques que nous formulerons, nous tenterons de proposer une solution de rechange.

Elster se donne comme objectif de fonder les normes sociales sur les émotions de honte et de mépris : "(...) social norms regulate behavior through the twin mechanisms of shame in the subject and disgust or contempt in the observer" (Elster 1999: 154-55). La culpabilité et la colère peuvent également entrer en jeu, mais de façon quelque peu secondaire; il est acquis ici que la culpabilité joue un rôle semblable mais moins important que la honte. Cette volonté

d'associer la honte aux normes sociales l'amènera à définir la honte comme une émotion strictement *sociale*, selon la typologie employée plus haut. C'est donc dire qu'il rejette son aspect personnel. Elster est clair là-dessus : "Other emotions, such as anger or shame, arise only when there is social interaction (...)" (141). Dans une note de bas de page (152, n.44), il admet la possibilité de honte personnelle à travers une interaction sociale imaginée, mais n'y accorde que peu d'intérêt. Il faut y voir selon moi une volonté de simplification du concept de honte afin de pouvoir l'intégrer dans son modèle des normes sociales. Mais comme il admet sans problèmes la possibilité de culpabilité personnelle (150-51), on voit mal pourquoi il refuse cette distinction à la honte. Selon Barrett (1995: 27), cette définition purement publique de la honte constitue une "approche traditionnelle" qui n'a plus la cote parmi les théoriciens contemporains de la psychologie[61].

Elster décrit la source de la honte comme suit : "Shame (...) arises when something one has done causes others to express disapproval. Because the disapproval takes the form of contempt or disgust rather than anger, it attaches to the person rather than to the act" (Elster 1999: 151). Et au sujet du sentiment de honte ou de culpabilité suivant la faute, "(...) one is forced into this attitude by the attitude of the observer" (151). Si la faute amène du mépris de la part d'autrui, cela causera de la honte; sinon, ce sera de la culpabilité. La culpabilité admet une cause supplémentaire, la violation d'une norme personnelle : "Feelings of guilt often occur when we have violated a *principle* that we view as binding" (150-1, italiques originales). Il semble que Elster se limite ici à l'auto-évaluation, excluant la possibilité qu'autrui force l'agent à évaluer son action sur la base de ses principes, mais cela demeure ambigu. Le modèle de Elster se résume ainsi :

a) La honte est causée par le mépris d'autrui, réel ou imaginé.
b) La culpabilité est causée (i) par la colère d'autrui, réelle ou imaginée, ou (ii) par la violation d'une norme personnelle.

Il est à noter que chez Elster, l'introspection revient avant tout à une interaction sociale virtualisée, du type "Si autrui m'avait vu agir, quelle aurait été sa réaction ?". Il n'y a pas de différences conceptuelles déterminantes entre ce type d'introspection et une critique réelle de la

[61] Pour un point de vue similaire, voir entre autres Lewis (1993: 569), Tangney et al. (1996: 1256-57) et Gibbard (1990: 137).

part d'autrui, seulement une différence de degré, la critique réelle étant par nature beaucoup plus influente que la critique imaginée.

Bien que Elster ait voulu élaborer un modèle concis des émotions et des normes sociales, il semble que celui-ci soit un peu trop simpliste. La relation directe entre le mépris d'autrui et le sentiment de honte (ou entre colère et culpabilité) ne peut se maintenir sans référence aux normes personnelles. Que l'exclamation d'autrui : "Vous êtes méprisable!", accompagnée d'une expression faciale appropriée, etc. cause de la honte chez l'agent, cela dépend du contexte, incluant les normes sociales intériorisées par cet agent. Si je me mouche bruyamment le nez sur la rue, et qu'un passant m'observe avec dégoût, je ne ressentirai de la honte que si j'entretiens de fortes normes de décorum (par exemple), de la culpabilité si ces normes sont plutôt faibles chez moi, ou rien du tout. La relation causale directe entre la réaction d'autrui et l'émotion, telle que préconisée par Elster, en est une hors contexte. Dans notre modèle, l'expérience de honte ou de culpabilité dépend des normes personnelles transgressées par l'action. Le mépris et la colère ne déterminent pas l'émotion, mais influencent le type de normes personnelles impliquées ainsi que l'intensité de l'émotion subséquente. Il est possible que le mépris affecte l'agent plus profondément que la colère, étant donné que le premier indique un rejet plus fondamental, alors que le dernier peut indiquer une volonté de poursuivre la relation ("qui aime bien châtie bien").

La relation entre choix rationnel et émotions, plus précisément entre action rationnelle et action "sous le coup de l'émotion", est notoirement complexe et, à ce moment, assez mal définie. Retournons à Elster pour voir comment il entrevoit cette relation. D'abord, l'émotion est une question de croyances : "(...) emotions are triggered by *beliefs* about events or states" (250, italiques originales). La relation croyance-émotion n'est ni purement logique, ni purement empirique mais bien de type "mécanique" (251), selon le concept des mécanismes sociaux cher à Elster. Il s'agit *grosso modo* d'expliquer une relation causale particulière à l'aide de principes suffisamment généraux pour qu'ils puissent servir dans d'autres situations semblables, sans viser la "grande théorie" universelle[62]. Les relations entre émotion et rationalité sont multiples.

---

[62] Voir à ce sujet le premier chapitre de Elster (1999) ainsi que le collectif dirigé par Hedström et Swedberg (1998).

Évidemment, l'émotion peut causer une action non rationnelle, mais Elster relève également que l'agent peut contrôler ses émotions jusqu'à un certain point, que celles-ci peuvent induire une irrationalité (par exemple en faussant les croyances), qu'elles peuvent être appropriées ou non à la situation, etc. A la fin de son chapitre sur la rationalité (328-31), il admet ne pas avoir trouvé de modèle précis d'interaction, et il nous offre en guise de conclusion une longue liste d'effets des émotions sur le comportement. Et à la toute fin de l'ouvrage, il résume ainsi sa position : "The role of emotions cannot be reduced to that of shaping the reward parameters for rational choice. It seems very likely that they also affect the ability to make rational choices within those parameters" (413). L'agent peut choisir entre l'action rationnelle et celle suggérée par l'émotion, mais la "force psychique", variable, de cette dernière fait pencher la balance en sa faveur. En accord avec la définition primaire, nous pouvons affirmer que la force psychique de la honte est significativement plus intense que celle de la culpabilité et par conséquent, cette dernière laisse une certaine place à l'action rationnelle que la honte ne peut permettre.

Du point de vue de la rationalité, l'aspect des émotions le plus intéressant pour nous est la *tendance à l'action*, c'est à dire les comportements induits par les émotions, en comparaison avec ceux préconisés par la rationalité. Nous avons déjà remarqué que la honte a la fuite et le retrait comme tendances à l'action, et que, pour la culpabilité, c'est l'admission de la faute, l'excuse et la réparation. Ces tendances concernent l'*expérience* de l'émotion; on s'intéresse peu à son *anticipation*, une condition amenant des tendances à l'action bien différentes[63]. Elster n'y consacre que quelques lignes sans vraiment élaborer, même s'il prétend que "the role of shame in decision-making depends on whether it is anticipated or experienced" (156). Cet énoncé est à mes yeux beaucoup plus capital qu'il ne semble le croire. L'anticipation de l'émotion est déterminante pour les normes sociales car c'est son mécanisme d'adhésion de loin le plus commun. Les normes ne sont pas des règles constamment violées et punies en conséquence. Leur violation doit être un phénomène exceptionnel, sinon on devrait remettre en cause la pertinence des normes en question. Je propose même d'en faire la distinction fondamentale de notre définition opérationnelle, que nous allons maintenant développer.

---

[63] Jacoby (1994: 5-6) inclut dans le rôle de l'anxiété cette possibilité d'anticipation de la honte.

Étant donné le caractère négatif des émotions à l'étude ici, les actions induites par leur anticipation visent à éviter leur actualisation. Supposons le modèle individuel désir-croyance suivant : l'agent a un désir D_X d'obtenir X, ainsi qu'une croyance C_X sur le plan d'action (ou la motivation, comme nous le verrons plus loin) approprié lui permettant d'atteindre X. Il entretient également la croyance C_E que ce plan d'action aura des chances de causer l'émotion E s'il est appliqué; c'est en d'autres termes son anticipation de l'émotion. L'agent anticipe E sur la base de l'atteinte de X par un plan d'action particulier. Lorsque E représente la honte, l'agent cherche à réviser soit le but X ou le plan d'action, ou simplement tout abandonner, selon ce qui est perçu comme la cause de E de façon à éviter E le plus possible; c'est une opération de conformisme. Dans le cas de la culpabilité, le but et le plan demeurent. Outre la révision, l'agent peut chercher à redécrire son action de façon à la rendre plus acceptable aux yeux de ses "juges", incluant lui-même. Le but de la redescription est toujours le même, éviter l'actualisation de E.

La force psychique de la honte étant particulièrement intense, elle laisse peu de place à l'action rationnelle, résultant en un conformisme face aux normes en vigueur. Si X risque de provoquer de la honte, la "désutilité" de l'émotion est tellement forte qu'il serait vain de la comparer à l'utilité de X, alors aussi bien l'abandonner. Nous ne pouvons donc pas parler de véritable désutilité, car l'effet de la honte se ressent au-delà de la réflexion rationnelle, il "affecte la possibilité de faire des choix rationnels", comme le souligne Elster. Les choses se passent différemment avec la culpabilité. Sa force psychique n'est pas aussi intense, par conséquent elle permet à la rationalité de jouer un plus grand rôle. Ici la désutilité de l'émotion, et par conséquent la possibilité d'un calcul d'utilité espérée, fait plus de sens. L'intensité psychique de la culpabilité force jusqu'à un certain point la redescription, mais permet la réflexion rationnelle car l'échec résultant en une culpabilité vécue n'est pas aussi critique que dans le cas de la honte. En d'autres termes, l'agent peut "vivre avec" la culpabilité, alors que la honte est beaucoup plus insupportable. Même si l'agent anticipe pleinement la possibilité de culpabilité, il peut échouer dans sa tentative d'évitement en évaluant mal les paramètres de la situation (en particulier les facteurs de risque) ou en agissant de manière irrationnelle. L'anticipation de culpabilité n'est en soi pas assez forte pour contrer les tentations (Sabini, Silver 1997: 8), alors que la honte semble posséder la force nécessaire. Mais comment la honte

peut-elle survenir lorsque anticipée ? Citant George Loewenstein, Elster prétend que nous avons tendance à sous-évaluer la possibilité de honte et, surtout, son impact dévastateur (Elster 1999: 156). Il faut bien comprendre qu'il ne s'agit pas ici d'une sous-évaluation de la désutilité de la honte, car ce qui définit la honte dans notre modèle est justement cette absence de réelle désutilité, mais bien de la reconnaissance de certaines conséquences comme pouvant générer de la honte plutôt que de la culpabilité, un embarras inoffensif, ou rien du tout.

3.3 Mécanismes d'anticipation

L'anticipation de la honte ou de la culpabilité face à une action violant les normes sociales va conduire l'agent qui désire éviter ces émotions désagréables à se conformer aux normes. Il y a d'abord le conformisme "pur", qui ne pose pas de difficultés particulières; c'est lorsque l'agent anticipe de la honte à accomplir X car ce but est inacceptable. Celui-ci va alors simplement abandonner sa volonté de X. Les cas les plus intéressants surviennent lorsque l'agent réussit à accomplir X même en anticipant une de ces émotions. Cela devient possible lorsque le jugement négatif porte non pas (ou non seulement) sur le but, mais sur la *manière* d'y arriver, donc sur le plan d'action. Nous nous pencherons ici sur les motivations à l'action. En particulier, lorsque l'atteinte de X par intérêt personnel se voit sanctionnée par les normes, l'agent peut accomplir son but en adoptant une motivation compatible avec celles-ci. Ce conformisme porte donc sur les moyens plutôt que sur la fin.

Elster (1999) propose un modèle général de transformation des motivations, consistant en un "entrant", la motivation originale, le "moteur" effectuant la transformation et le "sortant", la nouvelle motivation. Le moteur constitue une espèce de motivation de second ordre, faisant du sortant une "motivation motivée" (Elster 2004). On retrouve trois types de motivations, selon une division visiblement inspirée des classiques : l'intérêt, la passion et la raison. Les deux premières n'ont pas besoin de description; elles sont synonymes de rationalité et d'émotion. La raison a une forme particulière chez Elster, reposant surtout sur les thèses de Habermas. C'est une motivation qui adhère aux principes de la vérité propositionnelle, de la justesse normative (impartialité, désintéressement, absence de passions) et de la sincérité (Elster 1999: 337-8). Enfin, les transformations sont de deux natures. La *transmutation* est une transformation inconsciente ayant pour but de rendre l'action plus acceptable pour l'agent.

Elle émane du conflit entre deux des désirs fondamentaux déjà postulés, la promotion de son intérêt personnel et le maintien d'une image positive de soi, dans le cas où l'agent ne se conçoit pas comme quelqu'un qui ne fait que maximiser son intérêt, particulièrement en situations sociales[64] (369). La *fausse représentation* (*misrepresentation*) est une transformation consciente visant l'approbation d'autrui (332). Celle-ci vise à éviter la désapprobation et le mépris d'autrui, ainsi que les sanctions matérielles (401). La transmutation prend la forme d'une duperie de soi (*self-deception*), car on ne peut changer ses motivations de façon intentionnelle et instrumentale dans le but de se débarrasser de l'émotion. Si l'agent sait qu'il ne fait que se raconter des histoires, l'émotion perdurera. L'agent doit donc "graviter" inconsciemment vers la nouvelle motivation. La fausse représentation appelle l'hypocrisie, car la redescription vise autrui plutôt que soi-même.

C'est en combinant intérêt, passion et raison dans l'une ou l'autre de ces transformations possibles que Elster explore les "alchimies de l'esprit", comme l'indique le titre de son ouvrage. Mais toutes les combinaisons ne sont pas permises. D'abord, seule la passion peut servir de moteur à une transmutation, étant donné sa nature inconsciente (334). Les trois motivations peuvent servir de moteur à une fausse représentation, mais nous allons surtout nous attarder sur la passion. Nous nous pencherons brièvement sur l'intérêt comme moteur afin de montrer que l'agent peut être amené à changer ses motivations d'un point de vue purement rationnel. Le mécanisme propre au respect des normes sociales par anticipation prend la forme suivante : une transformation de l'intérêt en raison, médiée par la passion de la honte ou de la culpabilité. En d'autres termes, l'émotion nous pousse à abandonner nos visées personnelles et à épouser les normes sociales en vigueur. Comme la honte chez Elster ne peut être que sociale, il réserve celle-ci pour la fausse représentation et la culpabilité pour la transmutation[65]. Étant donné que pour nous, autant la honte que la culpabilité peut s'avérer personnelle ou sociale, nous devrons considérer que ces émotions peuvent servir de moteur autant à la transmutation qu'à la fausse représentation.

---

[64] Bien entendu, celui qui tient la satisfaction de ses intérêts égoïstes pour une norme personnelle forte ne passera pas par une telle transmutation.

[65] Son langage demeure toutefois hésitant : "To act on a motivation that the actor finds unacceptable is painful. To act on a motivation that other people condemn is also painful. Typically, perhaps, the former pain is that of guilt, the latter that of shame" (Elster 1999: 332, italiques rajoutées). Encore une fois, il relie timidement la honte au mépris d'autrui.

Elster adopte une notion restrictive et contextuelle de la raison; une raison "discursive" qui concerne le débat public. Pour lui, l'énoncé "In all public debates, all speakers represent themselves as being motivated by reason" est "close to a conceptual truth" (372). C'est plus qu'une norme sociale, la raison entrant dans la définition même de l'argumentation : "To say, in a public debate, 'We should choose policy A because it is good for me,' is to show a fundamental lack of understanding of what it *means* to offer an argument for something" (373, italiques originales). Si l'"intérêt" représente la maximisation de l'utilité personnelle, et la "raison" le discours impartial, quelle est alors la place des normes sociales ? A quelle motivation fait-on appel lorsque l'agent obéit à une norme sociale (ne pas jeter ses déchets dans la rue, par exemple) n'impliquant pas un débat public ? La solution la plus simple est de considérer la raison discursive comme une norme sociale parmi d'autres, et non comme une quasi-vérité conceptuelle. Ceci n'enlève rien au modèle de Elster et nous permet d'y inclure toutes sortes de normes. C'est ainsi que nous allons considérer la raison dans notre modèle. Ceci ne constitue aucunement une critique de la notion, mais seulement une généralisation du terme en "normes sociales" plutôt qu'en "normes régissant le discours". Sans doute ce mot "raison" devient-il inapproprié ici, nous le conservons afin de suivre le texte de Elster.

Les transmutations émanent du conflit entre le désir de promouvoir son intérêt personnel et le désir de satisfaire son estime de soi. Ici, l'agent anticipe de la honte ou de la culpabilité à obtenir X par intérêt personnel, et préférerait obtenir X d'une manière "raisonnable", compatible avec ses visées morales. Comme cela implique une duperie de soi, l'agent ne peut passer de l'un à l'autre intentionnellement. Dans le cas de la honte, la force de l'émotion fait en sorte que l'agent doit éliminer toute référence à son intérêt, et ainsi fonder sa recherche de X entièrement sur la raison. La moindre parcelle d'intérêt ramènera le sentiment de honte. Livet (2002: 95-113) nous offre une interprétation de ce type de transmutation. Son modèle général des émotions se fonde sur l'idée que celles-ci expriment un différentiel entre nos attentes et la réalité, et ainsi nous motivent à une révision de nos attitudes. Dans le cas de la transmutation, l'agent reconnaît son égoïsme dans sa quête de X. L'émotion anticipée de honte devrait le pousser à changer sa préférence pour X, mais selon la hiérarchie des révisions de Livet, c'est la plus difficile à accomplir. Alors, l'agent sera porté vers une révision de

dérivation plus facile, changera son plan d'action et cherchera à obtenir X sur une base raisonnable. Toutefois, la dérivation ne réglera pas le problème fondamental de l'agent, et une émotion d'angoisse perdurera jusqu'à ce qu'il révise sa préférence pour X, ou encore qu'il affirme sa volonté égoïste de X comme valeur fondamentale. Ici aussi, la duperie de soi doit être complète pour réussir, mais Livet demeure pessimiste quant aux chances de succès de l'agent.

La culpabilité étant une émotion moins forte, la transmutation correspondante va permettre le calcul rationnel. L'agent peut alors décider consciemment de faire une exception à sa règle morale et ainsi d'accomplir X par intérêt. Toutefois, celui-ci doit entretenir une *raison* sincère et non instrumentale justifiant l'exception, du genre "une fois n'est pas coutume". L'intérêt est alors transmuté en un amalgame intérêt-raison. Ce type de méta-règle en appelle directement à la maîtrise de soi; en effet, des violations trop fréquentes de ses règles de conduite démontrent un manque de maîtrise de soi et peut aisément déclencher de la honte. Le fondement de la méta-règle dans la honte empêche la régression infinie, car, si elle relevait de la culpabilité, l'agent pourrait alors faire une "exception à l'exception" du genre "une fois n'est pas coutume de transgresser une règle à répétition".

Dans la fausse représentation, la redescription vise à convaincre autrui que l'on respecte les normes sociales en jeu. Mais comme la honte concerne des normes fortement intériorisées et qu'elle entrave la possibilité de choix rationnel, il n'y a pas à proprement parler de fausse représentation motivée par la honte, car un tel mécanisme va immédiatement glisser vers une transmutation. L'agent qui anticipe de la honte doit non seulement redécrire ses motivations aux autres, mais également à soi-même. La notion de fausse représentation motivée par la honte demeure quand même utile, car la présence d'autrui va imposer des contraintes supplémentaires sur les types de raisons acceptables. Il en est tout autrement dans l'anticipation de culpabilité. L'agent peut alors inventer toutes sortes de raisons susceptibles de plaire à autrui, incluant les excuses et les justifications, sans nécessairement y croire lui-même. La force de la culpabilité réside alors dans l'acte même de la redescription : c'est ce qui induit l'agent à ne pas exprimer de motivations fondées sur son intérêt personnel. Donc, même si l'agent peut pleinement user de rationalité dans la recherche de raisons, c'est cette *nécessité* de

recherche de raisons qui est invoquée par l'émotion. La fausse représentation motivée par la culpabilité est le seul des quatre cas possibles où l'agent n'est pas obligé de croire en la validité de ses propres raisons.

Soulignons pour terminer la possibilité d'un respect rationnel des normes sociales selon le modèle des transformations des motivations. Il s'agit de la fausse représentation de l'intérêt en raison motivée cette fois-ci par l'intérêt. L'agent s'exprime en termes de raisons plutôt qu'en termes d'intérêt afin d'augmenter ses chances d'obtenir X; en d'autres mots, la raison comme motivation exprimée s'avère plus payante que l'intérêt. Relevons deux cas généraux. Le premier relève des sanctions matérielles : s'il existe des sanctions contre l'expression de ses intérêts personnels, l'agent va rationnellement adopter des raisons plus acceptables. Le second concerne les buts exigeant un effort collectif : lorsque l'agent a besoin de la collaboration d'autrui afin d'obtenir X, il sera fréquemment plus sage d'exprimer son désir de X en termes satisfaisants pour autrui. Dans cette forme de fausse représentation, la transformation de l'intérêt en raison n'est pas forcée par l'anticipation de l'émotion; elle constitue une stratégie de maximisation. Comme ces deux "moteurs" permettent des redescriptions rationnelles similaires, il sera difficile de distinguer leurs manifestations et nous devrons composer avec une certaine zone grise entre les deux.

3.4 Types de redescription

Nous venons de voir que, lorsque les normes sanctionnent la manière particulière par laquelle on atteint son but, on peut s'y conformer en reformulant cette manière, de façon sincère ou hypocrite. Toutefois, une redescription maladroite pouvant générer *elle-même* de la honte ou de la culpabilité, l'anticipation de ces émotions va en quelque sorte contraindre les redescriptions possibles (personne ne veut passer pour un hypocrite). Lorsque l'agent tente de se duper soi-même, une redescription inappropriée peut lui faire ressentir un manque de volonté, d'intégrité morale et de maîtrise de soi. Dans la duperie visant autrui, un échec peut amener la colère et le mépris de l'auditoire, et signaler à l'agent un manque de loyauté et de respect des autres. Dans cette section, nous allons tenter d'élaborer un inventaire des principales contraintes à la redescription ainsi que des diverses possibilités qui en émanent.

Dans les cas de transmutation, le critère le plus fondamental est la *véracité*. L'argument raisonnable doit être perçu comme vrai par l'agent. Il doit donc correspondre à sa perception du réel, car l'agent ne peut se mentir à lui-même intentionnellement et consciemment. Celui-ci est également soumis à la contrainte de *continuité* (*consistency*): la raison soutenue au présent doit demeurer compatible avec celles adoptées dans le passé à l'intérieur d'un domaine comparable (Elster 1999: 347-9). Cette contrainte émane de l'estime de soi : "The same need for self-esteem that caused us to justify self-interested behavior by impartial considerations in the first place may also prevent us from changing our conception of impartiality when it no longer works in our favor" (Elster 1999: 348). Même si la transmutation a pour objet l'agent, l'*environnement social* peut jouer un rôle déterminant dans la formation de motivations raisonnables. Une raison spécifique est plus aisément adoptée lorsque d'autres membres du groupe l'adoptent également, ceci renforce la confiance de l'agent envers sa rationalisation en lui conférant une certaine validité extérieure. Bien entendu, la raison de l'agent influencera celles des autres, créant ainsi un effet de groupe, un renforcement social de la duperie de soi (Statman 1997: 61). Ce mécanisme agit donc à la fois comme contrainte et comme incitatif.

Les contraintes à la rationalisation ne sont pas les mêmes lorsque nous passons aux cas interpersonnels de la fausse représentation. Plutôt que la véracité, l'argument doit ici respecter la contrainte plus faible de la *crédibilité*. La raison doit être crue vraie par autrui, mais pas nécessairement par l'agent (Elster 1998b: 103-5). Celui-ci aura alors la liberté de jouer sur le langage, le symbolisme et les rituels des normes, liberté absente dans la transmutation (Miceli, Castelfranchi 1998: 294). La véracité vient toutefois renforcer la crédibilité; plus un argument est sincère, plus il sera perçu comme crédible. La contrainte de *continuité* agit également dans la fausse représentation, mais au lieu de reposer sur l'estime de soi, elle vient à son tour appuyer la crédibilité. Se référant à Habermas, Elster (1994a: 215) soutient que la continuité est la forme extérieure de la véracité. C'est l'un des critères les plus utiles sur lequel autrui peut s'appuyer dans son jugement de la sincérité de l'agent, en autant que celui-là connaisse les prises de positions préalables de celui-ci. Puisque dans la fausse représentation (motivée par la culpabilité), l'intérêt demeure présent chez l'agent, celui-ci doit faire en sorte que sa raison ne paraisse trop intéressée. C'est la contrainte d'*imperfection* : une

concordance trop parfaite entre la raison exprimée et l'intérêt réel éveillera des soupçons chez autrui. L'agent doit donc démontrer publiquement qu'il abandonne une partie de son intérêt au profit du respect des normes, mais il ne peut non plus adopter une position trop désintéressée qui pourrait avoir des effets contre-productifs par rapport à son intérêt. Lorsque l'éventail des raisons possibles est restreint, il se peut que l'agent soit placé devant le douloureux choix d'une raison trop parfaite, donc peu crédible, et l'abandon de son intérêt (Elster 1999: 349, 376-7). Mentionnons enfin un mécanisme de passage de la fausse représentation vers la transmutation, même lorsque la culpabilité est en cause, ce que Elster nomme la "*force civilisatrice de l'hypocrisie*". Ce mécanisme se fonde sur un principe psychologique stipulant que l'hypocrisie génère des tensions désagréables chez l'individu (Elster 1999: 333; Statman 1997: 65). A force de répéter les mêmes raisons hypocrites (par la contrainte de continuité), l'agent en viendra naturellement à les croire sincèrement, ne serait-ce que pour soulager cette tension. Ce mécanisme se veut "civilisateur" car Elster croit que dans la délibération politique, le respect hypocrite de la norme d'impartialité se transformera en respect sincère, pour le plus grand bien de la collectivité (Elster 1994a: 248)[66].

La fausse représentation par la culpabilité permet à l'agent de tenter de leurrer autrui sur ses véritables motivations. Scott et Lyman (1968: 46) définissent le terme "*account*" (que l'on pourrait traduire par *motif* ou *explication)*, comme "(...) a statement made by a social actor to explain unanticipated or untoward behavior". Il existe selon eux deux types de motifs, les excuses et les justifications. L'excuse constitue un déni de responsabilité tout en admettant le caractère répréhensible de l'acte. La justification, au contraire, accepte la responsabilité tout en cherchant à minimiser ou à éliminer le caractère négatif de cet acte (Scott, Lyman 1968: 47). Ces types font partie des trois composantes nécessaires à la culpabilité, selon Miceli et Castelfranchi (1998: 295-9) : l'évaluation négative de l'action, l'admission de responsabilité et

[66] La transformation de l'intérêt en raison n'est qu'apparente, car l'intérêt particulier demeure, consciemment ou non, à l'esprit de l'agent. La force civilisatrice de l'hypocrisie propose une transformation réelle de l'intérêt en raison, le moteur de cette transformation étant l'habitude. Toutefois, Elster ne nous explique pas ce qu'il entend par habitude. Nous n'avons pas de description d'une situation générale où l'on pourrait s'attendre à ce que, par habitude, l'agent intériorise la norme qu'il respectait jusqu'alors de manière hypocrite. Il est évident que de telles mutations ont parfois lieu, mais une application générale de ce principe exige un mécanisme plus précis.

l'atteinte à l'estime de soi. Parmi les exemples d'excuses[67], nous retrouvons la difficulté de la tâche, la malchance, le manque d'information, la coercition, les circonstances atténuantes, la projection (attribution à autrui des mêmes motivations) et la sélection de causes lorsque plusieurs causes (d'agent ou d'événement) sont en jeu. Parmi les justifications, outre la diminution du caractère négatif de l'acte, retenons le refus de la critique (incompréhension, dérogation de ses juges), le refus de la norme et l'appel à une norme supérieure. Dans le passage de l'intérêt à la raison, l'agent conserve la responsabilité de son action; autrement dit, il ne prétend pas respecter les normes sociales par accident. C'est donc dire que la fausse représentation fera surtout usage de *justifications*. Comme la critique potentielle d'autrui concerne la violation des normes, la stratégie de l'agent sera de le convaincre que l'atteinte de X n'implique aucune violation ou du moins, que ses violations se justifient. Le besoin d'excuse survient principalement lorsque la critique porte directement sur X plutôt que sur la manière d'y arriver.

Les énoncés de motifs se déroulent dans un contexte social qu'il ne faut pas négliger. Il existe une série de normes régulant le genre de motifs appropriés à telle ou telle situation. Par exemple, le rabaissement du juge de ses actes ("Je me fous de ce que vous pensez!") peut être approprié face à une copine ou un parent, mais ne l'est certainement pas devant son patron ou l'électorat. Cette question est liée de près à la reconnaissance de l'autorité (Snyder 1985: 43), particulièrement dans les cas de justification, car on ne défie pas l'autorité lorsqu'on démontre que nous n'y sommes pour rien, comme dans le cas de l'excuse (Miceli, Castelfranchi 1998: 311). Scott et Lyman (1968) traite des motifs dans la théorie sociologique du rôle. C'est le rôle qui détermine le "style linguistique" adéquat du motif. Comme chaque individu occupe plusieurs rôles (parent, employé, membre d'une communauté ethnique, etc.), la stratégie des redescriptions peut alors s'étendre au choix du rôle. Enfin, l'expression de motifs acceptables permet la continuité des échanges sociaux, même dans les cas où l'auditoire *ne croit pas* en la sincérité de la justification de l'agent, pourvu que celle-ci respecte les normes régissant les motifs permis. Il s'agit d'une forme d'équilibre où chaque membre du groupe accepte de ne pas juger ses pairs à condition qu'ils ne le jugent pas en retour, dans les limites établies par les

[67] Ces exemples ont été tirées des typologies de Scott et Lyman (1968), de Snyder (1985) et de Miceli et Castelfranchi (1998).

normes. Au Parlement, chaque député sait bien que chacun redécrit fréquemment ses intérêts personnels (ou de son comté) en langage impartial, mais pourvu que ces raisons impartiales respectent les normes de délibération en vigueur, personne ne jugera personne. Cette réflexion de Pierre Bourdieu est éloquente : "Si ces tromperies qui ne trompent personne sont si facilement acceptées par les groupes, c'est qu'elles enferment une déclaration indubitable de respect pour la règle du groupe" (Bourdieu 1994: 233). La tromperie n'est pas entièrement cynique, car c'est la norme intériorisée qui oblige l'agent à décrire "correctement" ses intentions et selon Bourdieu, le groupe reconnaît cette obligation.

3.5 Conclusion

Dans la définition des normes sociales que nous avons retenue, l'agent se conformant à la norme d'une manière purement rationnelle ne sera pas perçu comme quelqu'un respectant l'essence de la norme. Afin de servir adéquatement la collectivité, la norme exige qu'on y laisse une partie de son intérêt particulier. Si l'agent ne se conforme qu'en apparence, cherchant des moyens acceptables pour promouvoir son intérêt, alors un modèle rationnel des normes demeure tout à fait approprié. Mais si l'agent se veut sincère dans son conformisme, s'il abandonne réellement la maximisation de son intérêt au nom des valeurs du groupe, alors le modèle rationnel sera déficient. En effet, comment modéliser rationnellement l'action d'un agent qui cherche consciemment à ne pas être rationnel ? Prétendre que celui-ci maximise l'intérêt du groupe au détriment du sien propre ne règle pas la question, car ce que la société demande, c'est le respect immédiat, irréfléchi, sans calcul des normes. Un modèle motivationnel combinant émotions et rationalité permet de combler cette lacune du choix rationnel en spécifiant les instances où une analyse rationaliste s'avère pertinente et, le cas échéant, en révélant certaines contraintes externes (véracité, continuité, etc.) limitant les choix.

En ouvrant la boîte noire du conformisme aux normes, nous découvrons un ensemble de comportements souvent ignoré ou trop superficiellement traité par le choix rationnel. L'apparence de conformité, dans le cas de la fausse représentation, requiert de l'agent des calculs stratégiques. Celui-ce devra choisir, parmi plusieurs raisons acceptables par le groupe, laquelle maximisera son utilité sous contrainte de crédibilité, de continuité et d'imperfection. Ce sont des calculs délicats et complexes, et si l'on rajoute le fait que certains agents en

position d'autorité ont leur mot à dire sur ce qu'est une raison acceptable, nous obtenons un champ d'application extrêmement intéressant pour le choix rationnel. Notre modèle des normes permet également une approche originale des phénomènes sociaux mettant en scène des agents interdépendants. A norme sociale donnée, on peut distinguer brièvement trois types d'agents selon leur motivation : celui qui tient la norme en haute estime et qui s'y conforme soit sincèrement, soit par transmutation de son intérêt; celui qui s'y conforme par solidarité au groupe sans vraiment s'intéresser au contenu de la norme; et celui qui s'y plie par pur calcul d'intérêt. L'étude de la dynamique d'un groupe composé de ces types de membres pourrait révéler, entre autres, à quel moment et en quels nombres chacun d'entre eux décidera de se conformer ou non[68].

L'intérêt principal de notre démarche est de permettre l'étude d'une classe particulière de comportement individuel se situant entre le non conséquentialisme et l'action rationnelle conséquentialiste. L'anticipation de honte ou de culpabilité motive l'agent à tenter d'éviter ces expériences douloureuses, et ainsi influence ses choix rationnels. De plus, la culpabilité nous permet d'agir de manière rationnelle, mais biaisée. Ces motivations et les mécanismes afférents, que nous avons voulu simples afin que notre démarche demeure compatible avec la théorie du choix rationnel, nous permettent d'analyser des phénomènes en-deçà du conformisme tels la redescription sincère, l'hypocrisie, la "langue de bois", le contournement stratégique des normes, etc. Nous avons introduit ici quelques notions clés pouvant servir d'outils à la théorie du choix rationnel. D'abord, nous avons tenté de définir deux nouveaux types d'agent : face à la transgression d'une norme, certains sont disposés à ressentir de la honte, d'autres de la culpabilité, avec des conséquences distinctes pour chacun. Nous avons également établi la différence entre l'anticipation et l'expérience de l'émotion en soulignant que cette première est en bien meilleure posture pour expliquer l'adhésion aux normes. Jusqu'à présent, cette distinction n'a pas été vraiment prise en considération dans la littérature. Finalement, nous avons brièvement esquissé une typologie des redescriptions possibles, présentée, encore une fois, dans un langage assez près de la rationalité.

---

[68] On retrouve un modèle dynamique de même nature, mais avec des type d'agents différents, dans Elster (1989b: ch. 5).

Le modèle que nous avons proposé n'est pas original. Il se veut une clarification ainsi qu'une généralisation du modèle déjà élaboré par Elster. Celui-ci propose de fonder le respect des normes sociales sur l'émotion de honte, laissant une place secondaire et quelque peu équivoque à la culpabilité. L'élément déclencheur de la honte est pour lui le mépris d'autrui, agissant de façon directe sur l'agent. Par conséquent, les mécanismes de transformation auxquels il fait appel ne distinguent pas bien les instances de honte et de culpabilité et demeurent à ce point de vue incomplets. En revanche, nous avons proposé une distinction et une catégorisation de ces émotions que nous croyons plus claire, en postulant que les normes se rattachent directement aux émotions à travers les valeurs de l'individu et indirectement aux sanctions externes, ce qui nous a permis d'élaborer, de manière explicite pour chacun, deux types de transmutations et de fausses représentations avec des conséquences distinctes pour l'action. Il est de notre avis que notre approche est plus systématique et permet l'analyse d'un plus large éventail de phénomènes que celle d'Elster, sans pour autant s'en éloigner substantiellement.

Certains champs de la vie sociale se prêtent plutôt bien à l'application de ce modèle. On pense immédiatement aux débats de nature politique au sens large, où certaines normes d'impartialité régissent le discours, ces normes variant selon que les discussions se déroulent à huis clos ou en public. L'analyse à la fois normative et rationnelle du comportement politique se situe à mi-chemin entre les théories du choix social et de la démocratie délibérative (cf. chapitre II). Il y a également les instances formelles de négociation telles les luttes syndicales ou certains types de transactions marchandes, où pouvoir formuler une concession sans perdre la face s'avère aussi, sinon plus, important que la concession elle-même. Un autre champ prometteur serait celui des relations hiérarchiques à l'intérieur d'organisations, où les ambitions personnelles se confrontent aux buts collectifs organisationnels. Il apparaît à travers ces quelques brefs exemples que notre modèle émotif-rationnel des normes sociales ouvre la voie à l'étude des comportements individuels *politiques* dans des domaines qui ont été amplement scrutés par d'autres méthodes mettant l'accent sur les résultats plutôt que sur le processus de décision.

CHAPITRE IV

LE CONFORMISME NON RATIONNEL AUX NORMES SOCIALES, SINCÈRE ET HYPOCRITE

Nous avons vu au premier chapitre que les modèles de choix rationnel aspirant à expliquer le phénomène du conformisme aux normes sociales se subdivisent en deux catégories : ceux qui s'en tiennent aux utilités standards, et ceux qui font appel à des utilités spéciales. Nous retrouvons dans la première catégorie les modèles dit d'"attentes mutuelles", dans lesquels les agents se conforment à la norme afin de réaliser un surplus collectif quelconque; sans la norme pour coordonner les attentes, la rationalité individuelle assurerait l'échec de l'entreprise. Nous retrouvons aussi dans cette catégorie les modèles dans lesquels le conformisme est motivé par la valeur intrinsèque de la norme en elle-même ou du sentiment d'appartenance au groupe. Au sein de la théorie du choix rationnel, de tels bénéfices au conformisme peuvent toujours être modélisés comme des préférences ordinaires, et des fonctions d'utilité peuvent sans problèmes être constituées autour de ces préférences. Toutefois, il arrive qu'à l'observation, les agents ne se comportent pas comme le suppose le modèle. Par exemple, le modèle peut stipuler que chaque agent recherche le surplus collectif, et on se rend compte que certains d'entre eux se conforment uniquement parce que les autres le font. Ceci ne constitue pas véritablement un problème dans la théorie classique du choix rationnel, car les préférences peuvent toujours être modélisées "comme si" tous recherchaient le surplus. Ce qui compte d'abord, c'est la plausibilité empirique des résultats du modèle plutôt que le réalisme psychologique des agents.

Certains modèles rationnels ont cherché à prendre en considération des motivations alternatives. L'agent peut suivre la norme à cause des valeurs qu'elle véhicule, ces valeurs étant considérées plus ou moins irréductibles aux préférences instrumentales. Les modèles de cette seconde catégorie ont recours à des utilités non standards, où l'utilité globale individuelle est fonction à la fois d'utilités instrumentales et non instrumentales. Ils font alors face à un dilemme de commensurabilité : en effet, soit que leur conception des utilités non standards n'est pas différente de nature des utilités standards, et alors ces utilités diverses peuvent s'intégrer sans heurts dans une seule fonction d'utilité, ou elles se révèlent incommensurables, et alors la

sans heurts dans une seule fonction d'utilité, ou elles se révèlent incommensurables, et alors la fonction d'utilité ne peut être constituée sans violer les postulats de la théorie de l'utilité et par le fait même, ceux de la théorie du choix rationnel.

La stratégie des utilités spéciales nous apparaît peu prometteuse. Nous maintenons cependant que les fondements motivationnels au conformisme non instrumental ne sont pas commensurables avec les préférences instrumentales typiques de la rationalité. La stratégie que nous proposons dans ce chapitre consiste à modéliser les motivations non instrumentales fondées sur les valeurs en dehors de la théorie du choix rationnel, tout en retenant les diverses possibilités de conformisme rationnel. Notre justification à soustraire à l'analyse purement rationnelle le phénomène du conformisme aux normes provient de deux sources. La première est le modèle des "motivations mixtes" de Elster (1999), qui distingue les motivations rationnelles des motivations émotionnelles pour les opposer. Ici, les émotions ne peuvent être modélisées comme de simples facteurs de la fonction totale d'utilité, car elles ne font pas qu'influencer l'utilité, elles peuvent aussi *fausser*, voire même *entraver* la réflexion rationnelle de l'agent. Dans notre reformulation de l'approche de Elster au chapitre précédent, nous avons proposé de considérer le conformisme non instrumental comme une tentative d'échapper à la honte ou à la culpabilité qu'entraînerait possiblement une violation de la norme. Nous avons également relié la honte aux valeurs fortes pour l'agent, et la culpabilité aux valeurs moins fortes. La honte est un sentiment négatif assez fort pour bloquer le raisonnement rationnel, alors que la culpabilité permet une certaine capacité de raisonnement rationnel, raisonnement qui aura toutefois un parti pris envers l'éradication du sentiment de culpabilité. Comme nous allons le voir plus en détails, l'anticipation de honte tendra à induire un conformisme sincère, alors que l'anticipation de culpabilité pourrait générer un conformisme hypocrite, c'est-à-dire un conformisme égoïste, publiquement affiché comme une promotion du bien de la communauté. Notre seconde source est une interprétation des idées de Bourdieu sur la reconnaissance collective. Principalement, il affirme que le groupe reconnaît et récompense davantage ses membres qui sacrifient leur propre intérêt au nom de l'intérêt collectif, par opposition à ceux qui, tout en contribuant autant à l'effort collectif, le font par égoïsme. Transposé en termes de choix rationnel, ce principe prend la forme suivante : *la manière la plus efficace de se conformer à la norme consiste à se comporter de façon non rationnelle.*

Même dans les cas où l'estime d'autrui est recherchée de manière intéressée et instrumentale, l'agent devra quand même se comporter comme s'il se souciait réellement du bien-être de la communauté. Les modèles rationnels d'attentes mutuelles sont mal outillés pour rendre compte des aspects symboliques et rituels des pratiques du conformisme. Ils peuvent difficilement expliquer comment il se fait que les agents, tout en se conformant à la norme tel que le prévoit le modèle, affichent publiquement une distance par rapport aux motivations rationnelles postulées. Ces aspects symboliques du conformisme peuvent grandement affecter l'allure que prendra le comportement du groupe, comme nous allons le voir plus loin à l'aide de notre concept d'"équilibre hypocrite".

Nous n'avons pas la prétention d'affirmer que les modèles rationnels ne sont d'aucune utilité à l'explication du conformisme aux normes sociales. La modélisation d'agents de manière à ce qu'ils exhibent une préférence pour le conformisme, ou qu'ils agissent "comme si" ils maximisaient leur utilité d'une manière quelconque demeure conceptuellement valide. Ce type de modèle nous apparaît toutefois incapable de prendre en compte les motivations non rationnelles de façon adéquate. Ceci constitue à nos yeux une limite explicative considérable. Bien que les modèles faisant usage de préférences "comme si" peuvent expliquer de manière satisfaisante des phénomènes "macro" d'équilibres et de persistance des normes, ceux-ci se révèlent impuissants à expliquer les pratiques actuelles qui se cachent derrière ces motivations instrumentales idéalisées de la théorie du choix rationnel. Nous sommes convaincus que les aspects psychologiques et symboliques du conformisme aux normes sociales sont déterminants, au point où ils peuvent produire des résultats collectifs que la rationalité n'aurait pu prévoir, et même dans les cas où l'explication rationnelle s'avère correcte, notre approche peut nous aider à comprendre pourquoi certains agents adoptent des pratiques visant à ne pas paraître rationnels. Notre but dans ce chapitre est d'examiner, du point de vue conceptuel, les effets d'un modèle de motivations mixtes sur l'action collective. La première partie sera dédiée à une élaboration conceptuelle des types de motivations individuelles au conformisme à la norme, en mettant l'accent tout particulièrement sur la reconnaissance sociale. Dans la seconde partie, nous entreprendrons une brève étude critique d'une approche rationnelle similaire à la nôtre, soit l'"économie de l'estime". Ensuite, nous tenterons de modéliser un équilibre normatif particulier, que nous nommons "équilibre hypocrite", dans lequel la plupart des agents

prétendent se conformer sincèrement à la norme dans le but d'obtenir l'estime d'autrui. Puis, nous appliquerons ce modèle à la pratique du duel. Enfin, nous terminerons ce chapitre en discutant des limites des modèles rationnels et des mérites de l'alternative que nous proposons. Avant de se lancer dans la modélisation, nous devons d'abord clarifier certains concepts tels l'agent, les normes sociales et la duperie de soi (*self-deception*).

Les agents de notre modèle sont supposés être motivés par les émotions et la rationalité. Nous avançons à cet effet quelques postulats fondamentaux. D'abord, nous distinguons préférences et valeurs. Alors qu'agir à l'encontre de nos préférences nous apporte une perte relative d'utilité, agir à l'encontre de nos valeurs nous conduit à une expérience de honte ou de culpabilité. Donc, chercher à éviter la honte ou la culpabilité motive en partie l'action au nom de ses valeurs[69]. Comme nous en avons discuté au chapitre I, ce type de motivation s'avère dans une certaine mesure non rationnel. Il pourrait se concevoir comme non instrumental : les valeurs génèrent potentiellement de la honte ou de la culpabilité précisément parce qu'elles sont perçues comme des fins en soi[70]. Il pourrait également se concevoir comme instrumental sans être pleinement rationnel : l'agent respecte ses valeurs dans le but (conscient et intentionnel) d'éviter la honte ou la culpabilité, mais son raisonnement se voit influencé par l'anticipation de ces émotions, en supposant que l'anticipation elle-même puisse provoquer une réaction émotive[71].

Nous concevons les normes sociales comme des règles externes et informelles, partagées par les membres du groupe approprié et maintenues par des sanctions. De plus, et là nous divergeons des définitions traditionnelles des normes sociales en choix rationnel, ces normes incorporent et expriment des valeurs partagées par les membres. Lorsque l'agent

---

[69] A ma connaissance, Elster n'affirme pas directement une telle thèse dans son modèle des motivations mixtes; ceci constitue une interprétation. Dans son modèle des transformations des motivations (1999, ch. V), la honte et la culpabilité jouent le rôle de "moteur" motivant l'agent à transformer ses motivations égoïstes en motivations "raisonnables" fondées sur certaines valeurs. Nous sommes d'accord avec Etzioni lorsqu'il affirme : "Values that have lost their affective elements become empty shells, fragments of intellectual tracts or phrases to which people pay lip service but do not heed much in their choices" (Etzioni 1988: 105).

[70] Rawls (1971: 440-46) établit un rapport entre les valeurs et le "bien premier" du respect de soi, et les émotions de honte et de culpabilité à la perte de ce respect de soi.

[71] Par exemple, l'anticipation de honte peut générer de l'anxiété (Jacoby 1994: 5-6).

endosse personnellement les valeurs exprimées par la norme, nous dirons que la norme est intériorisée par l'agent. Celui-ci aura désormais une motivation non rationnelle de s'y conformer. Nous postulons la caractéristique fondamentale suivante des normes sociales, qu'en tant qu'expression de valeurs, elles exigent toujours un certain sacrifice des intérêts égoïstes de l'agent en faveur des intérêts collectifs. Une telle conception s'éloigne sensiblement de la position de Elster[72]. Chez lui, le comportement non rationnel se veut principalement une conséquence des motivations émotionnelles sous-tendant les normes sociales. La violation de la norme entraîne le sentiment de honte ou de culpabilité, et l'anticipation de ce sentiment empêche l'agent de raisonner de façon pleinement rationnelle. Dans notre conception, la norme contient une injonction supplémentaire enjoignant l'agent à ne pas chercher à maximiser son utilité. Au lieu de prendre la forme "Tu dois faire X", nous obtenons "Tu dois faire X, et n'en profites pas pour maximiser ton utilité"[73]. Cette conception de la norme sociale s'oppose à la convention, soit une règle qu'il est acceptable de suivre d'une manière tout à fait rationnelle[74]. Celui qui se conforme à une norme sociale d'une manière ouvertement rationnelle subira les sanctions d'autrui, ou du moins, il n'en recevra pas autant d'estime qu'un conformiste désintéressé. Une règle informelle sera une norme sociale si elle incorpore certaines valeurs communes ainsi que certaines manières spécifiques de s'y conformer. Si aucune valeur n'est présente et que n'importe quelle manière légale de se conformer à la règle s'avère acceptable, nous parlerons d'une convention[75].

---

[72] Elster (1989b: 97-101; 1994b; 1999: 145-49).

[73] Même si Elster n'endosse pas directement une telle définition, il défend une position similaire comme contrainte au comportement, la "contrainte d'imperfection", qui stipule que lorsque l'agent fait face à un éventail de manières de se conformer à la norme. il devrait en choisir une qui ne correspond pas à ses intérêts de façon trop évidente, car cela pourrait paraître suspect aux yeux des autres (Elster 1999: 349, 376-77). Pour Elster, cette clause anti-rationnelle semble avoir simplement le statut d'un conseil, comment se comporter afin de paraître crédible. Nous considérons cette clause comme une caractéristique essentielle et universelle des normes sociales.

[74] Lorsqu'un conducteur respecte le code de la route (une convention), sa motivation à agir ainsi, que ce soit par calcul rationnel ou par civisme, n'a aucune importance pour les autres conducteurs; seul le geste compte. Quand une personne porte du noir à des funérailles (une norme sociale), les motivations ont de l'importance. Par exemple, elle ne pourrait pas déclarer qu'elle a choisi sa tenue en fonction d'une maximisation de ses chances d'obtenir une bonne part de l'héritage. Ici, son intérêt personnel doit être masqué ou supprimé.

[75] Quéré (1995) établit une distinction entre une conception des normes comme "manière d'agir" et comme "raison d'agir". La première se veut instrumentale, et concerne les normes comme attentes mutuelles. Le conformisme y est individuellement rationnel. La seconde est non instrumentale, et réfère à une obligation inhérente à la norme. Notre modèle prend ces deux conceptions en

Pour terminer, nous abordons la notion de "duperie de soi". Notre définition sera simple. La littérature sur ce sujet est vaste et complexe. Nous n'avons besoin ici que d'une conception "phénoménale" de l'influence de la duperie de soi sur le choix; nous pouvons nous passer de ses mécanismes psychologiques[76]. Nous proposons de définir la duperie de soi comme une évaluation biaisée de manière non intentionnelle d'une inférence incertaine, évaluation causée par l'anticipation de déplaisir face à une évaluation correcte. L'exigence d'incertitude rend la duperie de soi impossible lorsque les faits sous-tendant l'inférence s'avèrent irréfutables. Supposons que Paul est un grand amateur de sport. Il sait que certains athlètes de haut niveau ont été pris à faire usage de stéroïdes. Il sait aussi que les stéroïdes améliorent grandement les performances, et que beaucoup d'athlètes performent tout aussi bien que ceux qui se sont fait prendre. Partant de ces prémisses, l'inférence normale à faire serait que la plupart des athlètes de haut niveau font usage de cette drogue. Mais cette inférence demeure incertaine; il est toujours possible, quoiqu'improbable, que ces autres athlètes soient "propres". Anticipant (à un niveau semi-conscient) une grande déception s'il en venait à la conclusion que ses héros ne sont que des tricheurs, Paul va accorder, d'une manière non intentionnelle, plus d'importance aux manifestations du contraire, comme des témoignages d'athlètes jurant n'avoir jamais touché à cela. Il préserve de cette façon une vision confortable du monde du sport. Toutefois, lorsque l'inférence devient certaine, comme lorsqu'un athlète avoue sa tricherie, la duperie de soi ne peut plus opérer dans ce cas spécifique; et si suffisamment d'athlètes se joignent à la chorale, la duperie de soi cessera complètement de fonctionner. Comme les interactions sociales se veulent notoirement complexes et incertaines, nous postulons que le phénomène de la duperie de soi est fortement répandu, et donc un élément déterminant de notre modèle.

4.1 Les fondements motivationnels du conformisme aux normes sociales

---

considération; pour certains agents, la norme est une manière d'agir, pour d'autres, une raison d'agir.

[76] Pour Elster (1999: 362-63), il n'existe pas de définition adéquate de la duperie de soi au-delà de "métaphores". Voir Audi (1985), Lazar (1999), Scott-Kakures (1996) et Livet (2002: 95-113) pour des analyses détaillées pouvant s'appliquer à notre modèle.

L'approche des motivations mixtes nous fournit plusieurs pistes pour explorer le conformisme aux normes sociales. Nous avons choisi de porter notre attention sur une piste particulière, impliquant le désir de reconnaissance sociale. Les principaux désirs que l'agent cherche à satisfaire face aux normes sociales sont : l'estime d'autrui, l'estime de soi et, bien entendu, la maximisation de son utilité au sens large. C'est de cette manière que Elster établit les fondements de sa théorie, s'inspirant des moralistes français (Elster 1999: 85-94). Le désir pour l'estime d'autrui représente le désir fondamental d'appartenance et de reconnaissance sociale, et le désir pour l'estime de soi est une motivation à agir au nom de ses valeurs. Les deux peuvent générer de la honte ou de la culpabilité lorsque ces désirs sont frustrés; l'anticipation d'une telle frustration joue alors un grand rôle dans la prise de décision individuelle. L'estime d'autrui a besoin d'un agent externe pour sa satisfaction, alors que l'estime de soi est une affaire personnelle, quoique l'estime d'autrui puisse affecter l'estime de soi. Les agents se conformant à la norme de manière rationnelle ne sont pas préoccupés par les valeurs et les émotions afférentes; ils agissent en vue d'un gain matériel ou dans le but d'éviter les sanctions. Ils peuvent également se conformer dans le but d'obtenir la reconnaissance d'autrui, en autant que celle-ci représente un moyen pour une autre fin, comme le statut social ou les titres officiels. Les agents motivés avant tout par leurs valeurs sont supposés agir pour les satisfaire. Leurs motivations se veulent surtout émotionnelles, et ils se soucieront peu des récompenses plus matérielles ainsi que des menaces de sanctions, en autant que leurs effets sur la satisfaction personnelle ne sont pas trop déterminants.

Nous subdivisons le désir d'estime d'autrui en deux types, selon que la frustration du désir provoque de la honte ou de la culpabilité. Dans le premier cas, l'estime d'autrui est fortement reliée à l'estime de soi. L'agent s'auto-évalue à la lumière des jugements d'autrui sur sa personne, et le désir d'estime de soi exige que ces jugements soient *sincères*. Ce désir ne sera pas comblé si l'agent s'adonne à des manipulations dans le but d'induire les autres à lui accorder de l'estime. Dans le conformisme aux normes sociales, l'estime est idéalement octroyée aux agents qui ont contribué d'une manière altruiste au bien-être collectif, en accord avec les valeurs instanciées dans la norme. Toutes choses égales par ailleurs, plus le conformiste *paraît* être motivé par des valeurs plutôt que par son intérêt personnel, plus il recevra d'éloges, mais l'anticipation de honte fera en sorte que la jouissance de ces éloges

dépendra d'un respect sincère des valeurs. L'estime devient alors un produit contingent du conformisme fondé sur les valeurs : lorsqu'il est perçu comme étant fortement lié à l'estime de soi, le désir de reconnaissance sera satisfait seulement si l'agent ne le recherche pas intentionnellement. Si l'agent veut s'attirer les éloges d'autrui tout en anticipant de la honte à agir de manière si instrumentale, il peut toujours intérioriser la norme à travers la duperie de soi, développant ainsi une croyance que les valeurs de la norme ont de l'importance. Ce mécanisme ressemble à la "transmutation" de Elster, où une motivation inacceptable pour l'agent se voit transformée en une motivation plus acceptable à travers une motivation émotionnelle de second ordre (Elster 1999: 340-41). Ce processus n'est pas intentionnel; l'agent "dérive" vers la motivation acceptable, guidé par la satisfaction immédiate de son plaisir.

Le second type correspond aux cas où la frustration du désir pour l'estime d'autrui entraîne de la culpabilité. Comme le sentiment de culpabilité est moins intense que celui de honte, il affecte le raisonnement rationnel mais ne l'entrave pas comme a tendance à le faire la honte. Ici, l'agent peut intentionnellement rechercher l'estime d'autrui; toutefois, comme l'estime est octroyée avant tout aux conformistes motivés par les valeurs, l'agent se doit de donner une fausse impression de ses motivations réelles. Ceci correspond au second mécanisme transformateur de Elster, la "fausse représentation" (*misrepresentation*) : la motivation se transforme car elle se veut au départ inacceptable pour autrui. Cette représentation destinée à autrui peut résulter d'un calcul conscient et intentionnel; nul besoin ici de recourir au mécanisme non intentionnel de la duperie de soi. Toutefois, même si l'action hypocrite provient d'un raisonnement rationnel, l'effet de l'anticipation de la culpabilité est de *forcer* l'agent à adopter une posture hypocrite : on ne peut donc qualifier l'acte de pleinement rationnel[77]. La fausse représentation peut également être motivée par l'intérêt. Lorsque l'agent recherche le gain matériel, par exemple, le conformisme hypocrite peut s'avérer être le choix rationnel; toutefois, comme un tel agent n'est pas sous l'emprise des émotions, l'hypocrisie demeure pour lui un choix, et non une obligation.

---

[77] Kuran (1995: 4-5) remarque que même si la falsification de ses préférences peut s'avérer rationnelle, le fait de "vivre dans le mensonge" peut provoquer des émotions négatives telles que la culpabilité, la colère et le ressentiment, celles-ci pouvant affecter les actions futures. Le modèle qu'il propose ressemble à la fausse représentation motivée par la culpabilité, sauf que dans celui-ci, l'émotion naît d'une auto-évaluation de soi comme hypocrite, plutôt que de la crainte de se faire prendre à mentir.

L'attribution d'estime peut également être motivée par les émotions ou la rationalité. L'évaluateur peut afficher une satisfaction sincère et immédiate face à un conformiste, de la colère ou du mépris face à un déviant, ou encore approuver (ou désapprouver) les comportements d'autrui de manière à maximiser son utilité. Normalement, une approbation intéressée ne contribuera pas à satisfaire le besoin d'estime du sujet, mais comme la reconnaissance se veut intrinsèquement plaisante, une telle approbation peut être endossée par duperie de soi. Le sujet adopte alors de fausses croyances quant aux motivations de l'évaluateur, comme une redescription de son comportement comme n'étant pas vraiment intéressé, en autant que plane un certain degré d'incertitude sur la véritable nature de ses motivations. La désapprobation intéressée peut ne pas affecter le sujet. Si A sait que le mépris que B lui exprime n'est qu'une tentative délibérée de B d'empêcher A de profiter d'un bien commun, cela pourrait ne pas provoquer de honte ou de culpabilité[78]. Une autre perspective serait la suivante : "The need to belong is so strong that its protective monitor, that is, self-esteem, alerts one to perceived threats of exclusion in an automatic manner which does not depend on rational reflection, though such reflection might sometimes be able to turn off the alarm, so to say, if found to be ungrounded. Some diminishing of our self-esteem occurs automatically even when an insult comes from people who pose no real threat to our social status" (Statman 2000: 533). Une réaction rationnelle à une désapprobation rationnelle est toujours possible, mais nous devons retenir du commentaire de Statman que les émotions vives que déclenche parfois l'exclusion sociale peuvent entraver une estimation rationnelle de la situation.

Le coût de la sanction négative pour l'évaluateur peut servir de signal. Si le fait de sanctionner A représente une perte nette pour B, cela peut démontrer à A que la réprimande a une origine essentiellement non rationnelle : "The more it costs me to refuse to deal with you, the stronger you will feel the contempt behind my refusal and the more acute will be your shame" (Elster 1999: 146). L'évaluateur peut aussi s'avérer sensible à la reconnaissance d'une tierce partie de la performance de son rôle; par exemple, C exprime son approbation à B d'avoir adéquatement sanctionné A. Ceci équivaut à se conformer à une norme de second

[78] Quoique cela puisse provoquer de la colère ou de l'indignation (Elster 1999: 149).

ordre, reliée à la norme sociale initiale et précisant les bonnes manières de sanctionner les déviants. La motivation de l'évaluateur émane ici plus de l'anticipation de honte ou de culpabilité face au non respect de la norme de second ordre que d'un sentiment quelconque envers le sujet. En fait, un conformisme superficiel de la part du sujet serait probablement suffisant pour que l'évaluateur approuve celui-ci selon les recommandations de la norme de second ordre. Même si dans un tel cas l'évaluation n'est pas vraiment sincère, la reconnaissance du groupe envers l'évaluateur pourrait faire en sorte que le sujet ressente quand même l'évaluation comme si elle était sincère.

Le conformisme hypocrite doit respecter certaines contraintes. La principale d'entre-elles est la crédibilité : le conformisme instrumental doit mimer le plus possible le genre de comportement qu'aurait un agent motivé par les valeurs dans des circonstances similaires. Les comportements des agents sincères servent ainsi d'étalon à partir duquel les formes de conformisme seront évaluées. De plus, les conformistes sincères auront naturellement tendance à être aussi des évaluateurs sincères, établissant ainsi un second étalon, celui des sanctions appropriées[79]. Elster propose, dans la même veine, une contrainte de cohérence, stipulant qu'un agent ne devrait pas trop dévier de ses actes passés (Elster 1999: 347-49). Un conformiste sincère a un comportement habituellement assez stable quant aux instances répétées d'une même norme; les conformistes hypocrites devraient en prendre note. Et, bien entendu, puisque les normes sociales contiennent une injonction à ne pas agir de façon rationnelle, les conformistes hypocrites ne doivent pas trop paraître maximiser leurs intérêts. Lorsque l'on prend en considération les mécanismes de duperie de soi, d'autres contraintes font leur apparition, notamment le caractère plausible des interprétations faussées et la qualité de l'information.

4.2 L'économie de l'estime

Au-delà des motivations de nature émotionnelle, les agents ont intérêt, au sens rationnel du terme, à rechercher l'approbation d'autrui ou du moins à éviter leur

[79] Les évaluateurs sincères ne sanctionnent pas nécessairement toute forme d'hypocrisie. Seuls les plus radicaux d'entre-eux refuseraient inconditionnellement la contribution d'un conformiste hypocrite au bien-être collectif (un don de charité motivé par l'image, par exemple). Néanmoins, de tels évaluateurs auront tendance à faire preuve de plus de vigilance quant aux motivations des conformistes.

désapprobation, alors qu'il est dans l'intérêt des évaluateurs d'approuver ou de désapprouver les actions d'autrui selon les effets de ces actions sur leur propre utilité. Récemment, un modèle d'action collective nommé "économie de l'estime" tentait d'élaborer la manière dont des sujets et des évaluateurs pourraient se "marchander" en quelque sorte de l'estime. Du point de vue de la théorie du choix rationnel, le principal intérêt d'un tel modèle est qu'il résout le problème de second ordre des normes sociales en concevant l'estime comme une sanction, positive ou négative, qui ne coûte rien[80]. Un évaluateur actif n'est même pas nécessaire, car le sujet peut, par empathie, inférer ce que les autres pensent de lui, et ainsi en ressentir les émotions correspondantes (Pettit 1990: 739-40), bien que ces émotions soient moins intenses que celles provoquées par de véritables évaluations externes. Nous jetterons un coup d'oeil à deux conceptions de l'économie de l'estime, celle de McAdams (1997) et celle de Brennan et Pettit (2004).

McAdams propose une approche nettement rationnelle. Ici, l'estime fait partie de la fonction d'utilité au même titre que les coûts et les bénéfices matériels résultant du conformisme à la norme. De manière plus spécifique, l'agent a un incitatif à suivre la norme lorsque le coût potentiel de perte d'estime est plus grand que le coût du conformisme. L'attribution d'estime s'effectue de manière relative; on gagne de l'estime en contribuant plus que la moyenne au bien commun, ou en encourant un coût de participation plus élevé que la moyenne. Du côté des évaluateurs, l'attribution correcte d'estime envers les conformistes et les déviants confère en soi de l'estime. Dans ce jeu de l'estime, les joueurs sont, à la limite, motivés à la fois par l'espoir de devenir le seul à se conformer à la norme (le "héros"), avec gain maximal d'estime[81], et la crainte de devenir le seul déviant, avec perte maximale d'estime (McAdams 1997: 370-71). Ce modèle resserré suppose que l'agent se conforme à la norme de manière instrumentale, dans le but de gagner de l'estime ou d'éviter d'en perdre. Il n'interdit pas

[80] Le problème de premier ordre des normes sociales est : comment amener les agents à se conformer ? Typiquement, en choix rationnel, les solutions proposées mettent en jeu la promesse de récompenses et la menace de sanctions. Le problème de second ordre devient alors : comment amener les agents à sanctionner autrui, considérant que les sanctions sont coûteuses, et qu'il est donc préférable pour chacun de laisser les autres s'en occuper ?

[81] Le héros reçoit une très haute estime des autres qui, par définition, ne se conforment pas (encore) à la norme. Ce qui est ici estimé est le risque d'échec et d'humiliation que prend le héros agissant seul (McAdams 1997: 370). Je doute de la validité empirique de cette remarque. Les gens qui se dévouent au bien-être de la communauté sont souvent vénérés même dans la défaite.

les comportements ouvertement stratégiques dans le marchandage de l'estime. D'après Cowen (2002: 219), pour que l'"économie de l'estime" puisse fonctionner, les agents doivent, soit accorder une valeur en soi à l'estime intéressée, soit demeurer incertains quant à sa nature stratégique. Selon nous, l'estime intéressée n'a que peu de chances de satisfaire le désir d'estime, à moins que l'agent puisse glisser, par duperie de soi, vers une fausse croyance que l'estime reçue a été produite honnêtement, ce qui serait possible sous la seconde condition de Cowen.

Brennan et Pettit (2004) reconnaissent le caractère contre-productif du marchandage intentionnel d'estime[82]. Ils parviennent à maintenir l'idée d'une économie de l'estime en invoquant la possibilité d'un "échange virtuel", soit une disposition non intentionnelle envers un certain comportement qui sera tenu en estime (Brennan, Pettit 2004: 40-45). Ce comportement se maintient de manière fonctionnaliste par le sentiment positif causé par l'estime reçue. Ils supposent que la recherche d'estime rationnellement motivée est vouée à l'échec; par contre, ils postulent que le désir pour l'estime puisse *former* le comportement sans nécessairement le *motiver* (Brennan, Pettit 2004: 41). L'argument repose sur un contrefactuel : l'agent ne fait pas X dans le but d'obtenir de l'estime, mais advenant que X cesse de lui procurer de l'estime, il cesserait d'effectuer X[83]. Le désir pour l'estime sert de clause de réserve (*standby clause*) : il ne détermine pas le comportement de l'agent, mais dès que l'estime reçue tombe en-deçà d'un certain niveau de satisfaction, le désir ramènera l'agent dans le droit chemin (Pettit 1995). Comme la recherche ou la production intentionnelles d'estime est hors de question, ce modèle de l'économie de l'estime fonctionne alors par l'échange de "services d'estime" (*esteem-services*). Les agents adoptent des comportements à la périphérie de l'échange d'estime proprement dit, comme la sélection de ses partenaires, la présentation de soi devant un auditoire, ou encore le degré d'attention à accorder à une personne méritant notre estime. Ces stratégies fonctionnent en autant qu'elles ne soient pas entreprises explicitement dans le but de

---

[82] Voir aussi Elster (1983), ch. II.5, "Trying to impress".

[83] Bicchieri (1993) emploie un argument contrefactuel similaire afin de démontrer que le conformisme non instrumental à la norme (par habitude) se révèle instrumental au second degré : "If I ever wanted to be different, or if I expected others to do something different, I would probably overcome the force of the norm" (Bicchieri 1993: 231).

gagner de l'estime. Elles ne sont pas immédiatement rationnelles, quoique la "clause de réserve" du désir pour l'estime fera en sorte qu'elles seront ultimement rationnelles.

Cowen (2002: 218-19) souligne un point important concernant les implications de l'estime comme sanction sans coût d'application. Si l'attribution instrumentale d'estime s'avère bénéfique pour l'évaluateur et ne comporte aucun coût, on devrait assister à une poussée inflationniste. En effet, qu'est-ce qui empêcherait à la limite un agent d'accorder une quantité infinie d'estime, gagnant ainsi une utilité infiniment élevée ? De plus, si nous considérons l'estime comme étant attribuée à ceux qui contribuent plus que la moyenne, la compétition ainsi générée devrait contribuer grandement à l'inflation. Pour Cowen, plusieurs facteurs limitent l'attribution d'estime, comme la réduction de l'influence marginale de l'évaluateur au-delà d'un certain niveau (vénérer quelqu'un comme un dieu n'induira pas habituellement une satisfaction également divine chez l'agent), et le fait que les évaluateurs sincères ne sont pas motivés à accorder plus d'estime que ce qui est mérité. Ces deux contraintes sont pour nous intimement reliées. Comme le niveau correct d'estime à accorder est déterminé en observant les évaluateurs sincères, et que ce type d'estime n'est pas sujet à inflation, il s'en suit qu'une attribution stratégique (hypocrite) d'estime au-delà de la normale déterminée par l'attribution sincère aura l'air suspect. McAdams ne confronte pas ce problème, alors que chez Brennan et Pettit, l'attribution non intentionnelle d'estime permet d'éviter l'inflation.

Notre critique principale de l'économie de l'estime porte sur le fait qu'un tel modèle ne peut se concevoir comme une véritable économie. Ce qui fait que l'on ressente des émotions agréables à recevoir de l'estime est dû, entre autre, à l'attribution d'estime comme évaluation honnête et désintéressée de ce que l'on vaut. Comme nous venons de le constater, étant donné que l'estime est un bien désirable et que l'attribution d'estime peut offrir des avantages stratégiques, quelque chose comme un "marché" d'estime peut très bien émerger, mais pour qu'il puisse opérer efficacement, il ne peut pas être *perçu* comme un marché. De là les efforts nécessaires pour camoufler ce marché à l'aide d'euphémismes, d'idéologie et de duperie de soi. Une modélisation rationnelle stricte de l'économie de l'estime peut fort bien s'avérer adéquate, mais, à notre sens, un élément essentiel lui échappera toujours, soit les pratiques particulières nécessaires au bon fonctionnement de ce genre d'économie. Un tel modèle ne peut expliquer ces

pratiques car celles-ci sont non rationnelles, ou du moins elles sont faites pour paraître ainsi. La motivation rationnelle que ce genre de modèle pose par hypothèse constitue précisément la manière de *ne pas* se comporter lorsqu'il s'agit du conformisme aux normes sociales.

4.3 L'équilibre hypocrite

Dans la section 4.1, nous avons passé en revue des possibilités de motivations, pour les conformistes aux normes sociales et les évaluateurs du conformisme, impliquant la rationalité et les émotions. On peut dériver plusieurs types d'équilibres sociaux de ces motivations, selon la composition du groupe que nous modélisons. Par exemple, advenant que tous adhèrent aux valeurs contenues dans la norme, cela produirait un équilibre très stable où tous se conformerait à la norme, et où la réceptivité aux coûts du conformisme aurait tendance à demeurer très faible. Nous nous intéressons dans cette section à un type particulier d'équilibre, que nous nommons "équilibre hypocrite", dans lequel la plupart des agents prétendent se conformer sincèrement à la norme et approuver sincèrement le conformisme d'autrui, la motivation principale étant la recherche d'estime. Cet équilibre est dit hypocrite à la fois au sens individuel stipulant que l'agent n'endosse pas véritablement les valeurs de la norme, et au sens collectif stipulant que, si les agents en venaient à "prendre la norme au sérieux", sans considérations pour leur reconnaissance, l'équilibre ne pourrait tenir plus longtemps, et en bout de ligne la norme sociale disparaîtrait. Comme alternative à l'approche rationnelle de l'économie de l'estime, nous proposons une lecture de l'"économie des biens symboliques" de Bourdieu, ajustée à notre modèle motivationnel du conformisme aux normes sociales.

Le fondement de l'économie des biens symboliques se situe dans le désir universel pour la reconnaissance par le groupe. Les biens symboliques tels l'honneur, la réputation, etc., ont comme "racine anthropologique" un désir pour l'estime de soi (ou "amour propre") et un désir pour l'approbation d'autrui (Bourdieu 2003: 240). En société, le meilleur moyen d'obtenir l'estime d'autrui consiste à se départir de son propre intérêt en faveur de l'intérêt commun. Le comportement que Bourdieu qualifie de "désintéressé", la "soumission à l'universel", a la propriété d'être universellement reconnu et récompensé (Bourdieu 1994: 164-67). Toutefois, des agents motivés un tant soit peu par la rationalité peuvent s'"intéresser" à une telle

possibilité de récompense. Bourdieu aborde cette difficulté en assouplissant son principe universel de reconnaissance : "Il n'y a pas de société qui ne rende pas hommage à ceux qui lui rendent hommage en *ayant l'air* de refuser la loi de l'intérêt égoïste" (Bourdieu 1994: 182, italiques rajoutées). Cette précision rend possible le conformisme hypocrite comme comportement rationnel déguisé en comportement désintéressé. De plus, et c'est là l'élément clé du modèle de Bourdieu, les membres du groupe peuvent approuver le conformisme hypocrite *même s'ils ne sont pas dupes*, et ce pour deux raisons. D'abord, le simple effort entrepris par l'agent pour faire croire qu'il suit la norme peut représenter en soi une forme de respect pour le groupe[84] (Bourdieu 1994: 233-34). En second lieu, l'économie des biens symboliques repose sur un "tabou de l'explicitation", car révéler la véritable nature d'un échange symbolique comme marchandage rationnel aurait pour effet de détruire cet échange (Bourdieu 1994: 179). La "méconnaissance partagée", une forme de duperie de soi collective, est un état dans lequel "chacun sait - et ne veut pas savoir - que chacun sait - et ne veut pas savoir - la vérité de l'échange" (Bourdieu 2003: 278; voir aussi 1980: 191-92). Le premier mécanisme suggère que le contenu de la norme pourrait ne pas s'avérer décisif dans la décision de s'y conformer et de sanctionner, ces comportements pourraient alors se suffire à eux-mêmes pour la communauté. Le second mécanisme suggère que les membres qui se conforment à la norme par souci de reconnaissance plutôt que pour son contenu ne sont pas motivés à chercher à savoir si une telle pratique est chose commune chez les autres membres.

Un conformisme approprié à la norme amène la reconnaissance d'autrui; par "approprié" nous entendons une adhésion (apparemment) sincère aux valeurs de la norme. Prenons d'abord un groupe d'agents tirant bénéfice de la reconnaissance par une satisfaction de leur désir d'être estimé, mais seulement comme produit indirect et non voulu d'un conformisme sincère. Ces agents ne pourraient jouir d'une reconnaissance qui n'émane pas d'une appréciation honnête de leurs actions. Ils ressentiraient de la honte si on en venait à découvrir qu'ils sont principalement motivés par la reconnaissance, peut-être même lorsqu'eux-mêmes en

[84] C'est une forme de la maxime bien connue, "l'hypocrisie est un hommage que le vice rend à la vertu" (Bourdieu 1994: 236). Elster en offre une interprétation individuelle : "It is an important part of our self-image that we believe ourselves and want ourselves to be swayed by reason rather than by passion or interest; hence conscious hypocrisy in the realm of esteem has a parallel in unconscious self-deception in the realm of self-esteem" (Elster 1999: 91). Chez Bourdieu, l'hommage porte surtout vers le collectif.

prennent conscience. De tels agents auront donc une motivation à transmuter leur désir de reconnaissance en désir de respect des valeurs de la norme, à travers une duperie de soi ayant comme résultat l'intériorisation de ces valeurs qui étaient auparavant extérieures. Ceux qui réussissent à ainsi intérioriser la norme pourront donc obtenir une reconnaissance dont ils sauront jouir. A la limite, lorsque tous les membres du groupe sont de tels conformistes "transmutés", un équilibre normatif peut apparaître dans lequel tout le monde est motivé à la fois par les valeurs et la reconnaissance, mais dans lequel tout le monde croit que tous les autres sont des conformistes uniquement mus par les valeurs. Kuran (1995: 81-82) maintient que les gens sont enclins à une "erreur d'attribution fondamentale" dans la prise de décision collective. Lorsque ceux-ci évaluent les décisions des autres, ils ont tendance à sur-estimer les facteurs intentionnels et subjectifs et à sous-estimer les facteurs externes comme l'influence du groupe. La duperie de soi et l'incertitude quant aux motivations d'autrui que nous postulons s'accordent avec un tel mécanisme. Si je me plie à la norme d'une certaine manière, résultat d'une motivation transmutée, jugeant donc ceci la "bonne manière" d'agir, j'aurai tendance à expliquer des comportements similaires chez les autres comme des instances de conformisme sincère, et peut-être approuverai-je en conséquence. Nous pourrions nommer un tel équilibre "semi-hypocrite", car les agents croient vraiment, quoiqu'à travers une duperie de soi, qu'ils se conforment de manière appropriée. Cet équilibre est en réalité plus faible qu'il apparaît par l'observation des comportements, car l'engagement "réel" envers la norme se veut moins fort que ses manifestations publiques pourraient nous laisser croire.

L'équilibre s'affaiblit un peu plus lorsque nous y introduisons la possibilité de ressentir de la culpabilité. Les agents motivés par la culpabilité ne ressentent pas d'attachements particuliers aux valeurs de la norme, et n'ont pas de difficultés à prétendre à l'attachement sincère dans le but de satisfaire leur besoin d'être estimé. En supposant que conformisme hypocrite et conformisme sincère se présentent de l'extérieur de façon identique, le groupe apparaît à l'observation soutenir fortement la norme, mais plus il y a de conformistes hypocrites, moins la norme jouit d'une base réelle solide. Les agents rationnels, non motivés par les émotions, peuvent se joindre au mouvement lorsque la reconnaissance devient source de pouvoir, en permettant d'accéder à des ressources supplémentaires ou à une position de domination au sein du groupe, par exemple. Eux aussi feront semblant d'adhérer aux valeurs

de la norme de manière désintéressée, et auront tendance à sanctionner les autres comme le ferait tout bon conformiste sincère[85]. Avec le temps, les agents moins sincères pourraient former le groupe entier des conformistes, menant ainsi à un équilibre pleinement hypocrite dans lequel tous paraissent suivre sincèrement la norme, alors que personne n'y croit vraiment. La possibilité d'un équilibre hypocrite dépend de quatre postulats concernant les normes sociales et la reconnaissance:

a) L'abnégation de soi en faveur du groupe se voit toujours récompensée.

Plus précisément, à contribution égale, une contribution désintéressée au bien-être du groupe sera plus appréciée par les membres qu'une contribution intéressée. Dans la détermination du degré de reconnaissance, les motivations des agents ont de l'importance, pas seulement leurs actions.

b) Les normes sociales contiennent des valeurs que les membres désirent pour leur communauté.

Ceci amène deux raisons différentes de se conformer à la norme : la rationalité, dans le but de maximiser sa propre utilité ou encore, sa conception de l'utilité collective, et les valeurs, lorsque celles de l'agent correspondent à celles de la norme. Ce dernier type de conformisme est avant tout stimulé par les émotions et se veut non instrumental; en d'autres termes, les coûts et les bénéfices du conformisme, autres que la satisfaction de son besoin d'estime, ne font pas idéalement partie du processus de décision. Comme le conformisme motivé par les valeurs se veut désintéressé et sera probablement à l'avantage du groupe, la norme sociale pourrait prendre la forme de la règle "officielle", quoique informelle, de la participation désintéressée.

c) *La reconnaissance est un bien recherché.*

Le conformisme génère de la reconnaissance, et celle-ci est toujours bénéfique pour l'agent, même si l'évaluation subjective de l'agent de cette reconnaissance peut varier. Pour les conformistes sincères, la reconnaissance est un produit non intentionnel de l'action, elle n'est jamais recherchée pour elle-même; elle peut toutefois faire l'objet d'une duperie de soi. La

---

[85] "(...) observance of a norm itself leads a person to impose the norm on others - even when the person himself observes the norm only because of the sanctions it carries" (Coleman 1987: 143).

reconnaissance peut être recherchée pour elle-même par des agents motivés par l'estime, et elle peut être recherchée de manière instrumentale, comme source de pouvoir, par des agents rationnels.

d) Un marché explicite de la reconnaissance s'annihilerait lui-même.

La reconnaissance (ou l'estime) constitue un bien spécial qui, de par sa nature même, ne peut pas s'échanger ouvertement d'après un modèle classique de marché. La recherche intentionnelle de reconnaissance provoquera probablement des réactions négatives de la part des conformistes sincères. L'approbation intéressée ne profitera pas à l'agent visé, à moins qu'il ne sombre dans une duperie de soi. La désapprobation intéressée peut ne pas affecter son récipiendaire. Même si une certaine "économie de l'estime" peut être inférée de l'observation des relations du groupe, elle ne constituera pas une description correcte du comportement individuel.

Des deux premiers postulats, nous en déduisons que les conformistes "authentiques", motivés par les valeurs, dicteront les manières appropriées de suivre la norme et d'appliquer les sanctions lorsque nécessaire. En ajoutant le troisième postulat, l'agent rationnel recherchant la reconnaissance aura de meilleures chances de succès en imitant les conformistes authentiques, autrement dit, en prétendant être motivés par les valeurs de la norme. Le quatrième postulat fournit à l'agent rationnel un incitatif à maintenir l'illusion de conformisme sincère, tant que tous croient qu'il existe un certain niveau de conformisme sincère au sein du groupe; combiné avec le troisième postulat, il fournit un incitatif supplémentaire à ne pas chercher à connaître le statut réel du conformisme des autres. Il s'agit ici d'une duperie de soi dirigée vers les motivations d'autrui, plutôt que d'une duperie de soi de ses propres motivations, présente chez les conformistes sincères attirés par la reconnaissance. Les phénomènes sociaux contribuant à maintenir la "méconnaissance partagée" seront également à l'oeuvre dans le maintien de l'équilibre hypocrite. Nous avons déjà fait connaissance avec l'"erreur d'attribution fondamentale" de Kuran; nous pourrions rajouter son mécanisme de falsification des préférences, mettant en scène des préférences publiques hypocrites, pouvant conduire à une ignorance répandue des aspects négatifs du *statu quo* (Kuran 1995: 19). Les membres d'un groupe peuvent s'influencer entre eux concernant leurs croyances émanant d'une duperie de soi

(Statman 1997: 61) : on glisse beaucoup plus facilement vers une fausse croyance lorsqu'elle est partagée par les autres, ici la croyance qu'il y a parmi nous des conformistes sincères, surtout si chacun croit que les autres ne sont pas sujets à une duperie de soi. La liste des mécanismes sociaux de type idéologique pouvant s'appliquer ici est longue. Ce qui est important à retenir, c'est que l'idéologie joue sur l'incertitude collective à propos des croyances et des motivations de chacun, et renforce la duperie de soi nécessaire au maintien de bien des normes sociales.

Nous retenons deux propriétés spécifiques de l'équilibre hypocrite, par opposition à l'équilibre normatif "sincère". La première est la possibilité d'un équilibre globalement néfaste pour le groupe. Un équilibre reposant avant tout sur la reconnaissance, et dans lequel les agents ne portent pas particulièrement attention aux valeurs de la norme ainsi qu'aux conséquences collectives de leur conformisme, pourrait maintenir une norme néfaste en autant que le bénéfice individuel de la reconnaissance demeure supérieur à la part individuelle des coûts globaux. La seconde est la faiblesse relative : un équilibre hypocrite s'avère intrinsèquement plus faible qu'un équilibre sincère, car il dépend d'une méconnaissance partagée pour se maintenir. Il manque à un tel équilibre le fondement axiologique caractéristique de l'équilibre sincère; néanmoins, les agents vivent dans l'illusion de sa présence. Ceci a pour effet de rendre les pratiques de conformisme plus embrouillées, et les règles plus complexes et voilées de symbolisme. En résumé, l'équilibre hypocrite exige une structure idéologique assez étendue, afin de masquer le fait que ses participants vivent des relations se rapprochant d'une "économie de l'estime" en bonne et due forme.

4.4 Une application

Si le conformisme à une norme sociale se veut idéalement non rationnel, alors le duel doit bien être la norme sociale par excellence : qu'est-ce qui est moins rationnel que de risquer sa vie simplement pour obtenir la reconnaissance de sa communauté ? Nous retrouvons dans les pratiques du duel plusieurs principes et mécanismes mentionnés plus haut. En France, le duel a pris naissance au Moyen âge, dans la double pratique du tournoi sportif où les combattants démontraient leur courage, et le "duel judiciaire", où les parties en litige se livraient un combat à armes égales, laissant Dieu décider de l'issue. A partir du XVe siècle, la

monarchie ainsi que l'Église catholique vont graduellement proscrire ces pratiques; toutefois, la classe aristocratique va perpétuer le duel, s'en servant comme d'un signe distinctif et comme d'une arène où l'on peut mériter de l'honneur (Billacois 1986: 31-40). L'honneur, et par lui l'appartenance à l'aristocratie, était très en demande à cette époque. Il menait souvent à des postes élevés dans les sphères de l'État et de l'armée. Le conformisme aux normes du duel exhibe un aspect axiologique pour ceux qui désirent vivre une vie honorable, et un aspect rationnel pour ceux recherchant l'honneur pour ses récompenses. Tout au long de l'histoire, les duellistes ont toujours eu soin de ne pas concevoir leurs pratiques dans un langage rationnel. Le "point d'honneur" que les gentilshommes avaient à défendre en tout temps représentait pour eux "l'essence même de l'honneur le plus pur, le plus raffiné et le plus exigeant" (Billacois 1986: 346). Aux non initiés, tout cela paraissait ridicule[86].

L'honneur ne s'acquiert ici que par des actions non rationnelles. Toute tentative de maximiser ses chances de survie dans un duel était considérée comme déshonorable. Les duellistes enlevaient leurs survêtements avant le combat, afin de bien montrer qu'ils ne portaient pas de plastrons. Des seconds étaient présents afin de s'assurer que les armes, la posture, etc. étaient équivalentes pour les deux adversaires. A l'ère de l'épée, les gentilshommes rationnels, s'ils voulaient améliorer leurs chances de victoire, devaient s'entraîner en secret (Elster 1999: 223). Celui qui avait le choix des armes prenait souvent celle dont il croyait que les autres croyaient qu'il ne possédait pas d'habiletés particulières à son maniement. Le pistolet a éventuellement remplacé l'épée, réduisant ainsi l'avantage de l'habileté. Dans un autre ordre d'idées, les combattants devaient démontrer qu'ils ne voyaient pas le duel comme un moyen d'éliminer l'adversaire : "(...) any kind of instrumental concern was dishonorable. The purpose of the duel is not to win, but to expose oneself to danger" (Elster 1999: 222). Le vainqueur épargne la vie de son adversaire avec générosité, il fait "don à un égal"[87] (Billacois 1986: 363). On retrouve également des motivations non rationnelles dans les affronts menant au duel. Un

---

[86] La classification de Bourdieu des pratiques sociales comme "jeux" colle de près à ce genre d'analyse. Ceux qui sont "pris au jeu" exhibent un grand intérêt pour des pratiques qui semblent futiles de l'extérieur. De plus, ceux-ci ne se perçoivent jamais comme participants à un jeu (Bourdieu 1994: 151-52).

[87] Billacois ajoute que les rois faisaient grâce comme don intéressé, obligeant ainsi leurs victimes. Épargner la vie d'autrui de manière désintéressée peut dont être vu à la fois comme une action non instrumentale et un geste distinctif, instrumental à un autre niveau.

gentilhomme pouvait refuser un défi ouvertement rationnel, comme celui provenant d'un homme de rang inférieur. Les duels les plus frivoles étaient ceux qui conféraient le plus d'estime. Un défi lancé à propos d'une insulte grave peut s'avérer la chose rationnelle à faire, au-delà de la recherche d'estime, dans le but de donner une bonne leçon à l'effronté et le persuader de ne pas recommencer. Dans les insultes plus frivoles, par contre, il y avait bien peu d'incitatifs à la réaction rationnelle. Toutefois, les défis ne devaient pas avoir l'air trop frivoles, sinon il deviendrait évident que le provocateur recherche n'importe quelle opportunité à se battre pour rehausser son statut (Hardin 1995: 93, 99). Cette possibilité était connue des duellistes d'expérience; ceux qui s'affrontaient pour le prestige s'adonnaient à des rituels étriqués, à cheval sur la mince frontière entre en faire trop et ne pas en faire assez pour suivre la norme. Comme contrainte supplémentaire à l'engagement au duel, l'agent honorable se devait d'apparaître intentionnellement désintéressé; la colère et la vengeance "à chaud" ne représentaient pas des motivations acceptables[88]. Cela se reflétait dans les règles du duel, comme les délais entre l'affront et le combat[89], la gestuelle polie avant le combat, l'arrêt des hostilités dès que coule le sang, et la réconciliation après le combat (Billacois 1986: 358-61).

Les normes de duel sont collectivement néfastes, au sens d'un sous-optimum de Pareto, car il existe des alternatives moins violentes permettant de régler le prestige et le rang dans les société dont nous avons traité. Ceci a pour effet d'affaiblir l'équilibre normatif, ce qui, selon nous, explique pourquoi le "code d'honneur" se présentait d'une manière beaucoup plus complexe que les normes sociales typiques, même pour l'époque. Il est frappant de constater qu'avec le temps, alors que l'on dénonçait la pratique de toutes parts, le code écrit gagnait en popularité, tandis qu'au début, le code était essentiellement oral, et sujet à une myriade d'interprétations (Billacois 1986: 313). L'explication que nous offrons à partir de notre modèle est que, à mesure que l'équilibre normatif s'affaiblissait, les "règles du jeu" au sens de Bourdieu devaient être resserrées afin de prévenir la faillite de l'équilibre. Durant l'époque des duels, des

---

[88] Bien entendu, les émotions liées aux normes, comme la honte, la culpabilité, le mépris, etc., étaient permises. Accepter un défi par peur de la honte en cas de refus est toujours intentionnel, même si une telle action n'est pas entièrement rationnelle.

[89] La règle XIV du Code irlandais du duel se lit comme suit : "Challenges are never to be delivered at night, unless the party to be challenged intends leaving the place of offence before morning; for it is desirable to avoid all hot-headed proceedings" (cité dans Halliday 1999: 179). Avant leur codification, les délais étaient parfois interprétés comme de la couardise (Billacois 1986: 398).

lois anti-duels étaient en vigueur partout. Néanmoins, il aura fallu des événements marquants, plutôt que des lois, pour mettre fin aux duels. Ceci est compatible avec notre modèle de duperie de soi : tant que les aristocrates (et leurs admirateurs) pouvaient perpétuer la croyance que le duel était essentiel à l'ordre social, la sphère judiciaire n'avait que peu d'influence sur eux[90]. Comparons la disparition de la norme du duel au Nord et au Sud des États-Unis au XIXe siècle (Wells 2001). Au Nord, le duel était une affaire plutôt rationnelle. C'était l'apanage des politiciens voulant prouver à la population qu'ils endossaient pleinement leurs convictions politiques, qu'ils n'étaient pas de ceux qui retournent leur veste au moment opportun. Conséquemment, la duperie de soi nécessaire au maintien de la norme s'avérait plus difficile à soutenir. Le célèbre duel de 1804 opposant Burr et Hamilton, dans lequel ce dernier trouva la mort, choqua la population, et la pratique fut abandonnée quelques années après. Elle survécut plus longuement au Sud. Là-bas, l'honneur occupe une place importante dans la culture, par conséquent, les valeurs de la norme du duel y sont plus fortement intériorisées. Les duels y étaient plus frivoles; ils y ont proliféré malgré le fait que tous les États de la Confédération l'interdisaient. Il aura fallu un événement beaucoup plus marquant, la Guerre civile, pour finalement mettre un terme à la pratique. Dans l'ensemble, la naissance du capitalisme et la montée de la richesse et de la réussite en affaires comme marqueurs de rang en société ont fourni cette alternative non létale mentionnée plus haut, rendant la norme du duel futile à peu près partout à la même période.

4.5 Conclusion

Le concept d'équilibre hypocrite que nous proposons nous a aidé à expliquer pourquoi des agents rationnels voudraient s'engager dans des comportements aussi compliqués que la participation au duel. La norme du duel instancie la valeur de l'honneur, une valeur largement partagée par les membres de l'aristocratie. Le conformisme intéressé à la norme est considéré déshonorant; seul le conformisme désintéressé est permis. Quelques agents sont motivés par l'honneur (et la crainte de la honte par le déshonneur), et ils agiront de façon non rationnelle, liés par l'honneur à s'engager dans des duels lorsque la situation l'exige. Ces conformistes sincères vont définir la façon correcte de respecter la norme, conférant de l'estime à ceux qui

[90] Nous devons ajouter que les branches législatives et exécutives de l'époque étaient peuplées d'individus qui, pour la plupart, souscrivaient à l'éthique du duel.

agissent de la sorte. Les agents qui ne sont pas (ou peu) motivés par l'honneur auront ainsi une raison de se conformer, une raison toutefois hypocrite. Enfin, si tous, ou la plupart, s'avèrent être des conformistes hypocrites, ils auront encore un incitatif à maintenir l'illusion et ne pas chercher à savoir la vérité, car s'ils découvrent qu'ils se servent de l'honneur de façon instrumentale, l'estime cessera d'être accordée. Une explication dans le paradigme du choix rationnel porterait son regard sur les bénéfices collectifs de la norme, et les mécanismes de maintien de ces bénéfices par le conformisme. Le modèle avancé par Schwartz, Baxter et RRyan (1984) tente d'intégrer des motivations non rationnelles dans un contexte rationnel. Leur conception des valeurs se lit ainsi : "values, which are embraced for noninstrumental reasons, may be important elements in the efficient realization of goals precisely because they are *not* viewed as justified by the consequences when they are adhered to" (Schwartz, Baxter, Ryan 1984: 331, italiques originales). La valeur de l'honneur tient dans leur modèle le rôle suivant : "A self-enforced sanction to behave honorably in these instances would raise the level of compliance and thus the social surplus available through adherence to the norms implied by the concept of honor" (Schwartz, Baxter, Ryan 1984: 348). Mais si l'on admet la présence de comportements *essentiellement*[91] non rationnels, pourquoi tenter de sauver la théorie du choix rationnel en postulant que l'honneur, en quelque sorte, s'avère instrumental à la production du surplus collectif ? Tout au long de leur article, les auteurs s'étonnent de découvrir des pratiques qui ne semblent pas s'accorder avec la rationalité. Nous proposons de prendre au sérieux ces comportements non rationnels et d'en étudier l'impact parallèlement à ceux des comportements rationnels.

D'un point de vue plus général, nous offrons deux critiques principales aux modèles de choix rationnel du conformisme aux normes sociales. La première est que leurs explications présentent bien souvent une tendance fonctionnaliste. Lorsqu'ils doivent confronter un comportement qui ne semble pas rationnel *prima facie*, les modèles rationnels ont tendance à souligner les bénéfices collectifs de la norme, et ensuite à attribuer rétroactivement aux agents la volonté de maximiser ces bénéfices. Ces modèles n'ont alors d'autre choix que de postuler

---

[91] Contrairement au comportement *apparemment* non rationnel, une approche beaucoup plus commune en choix rationnel. Dans ce cas, l'hypothèse de rationalité n'est pas rejetée, et la solution consiste à rechercher des préférences ou des facteurs de fonction d'utilité alternatifs.

des préférences et des motivations *ad hoc*, de type "comme si"[92]. Tant que le modèle vise l'explication macro des interactions sociales[93], ceci ne constitue pas une objection sérieuse : tous les modèles fondés sur l'individualisme méthodologique doivent idéaliser les comportements d'une manière ou d'une autre. L'objection prend son sens lorsque le modèle tente d'expliquer les pratiques de conformisme en elles-mêmes, comme nous venons de le voir dans le modèle du duel ci-haut. La seconde critique concerne l'aspect normatif de la théorie du choix rationnel. Le rôle de la théorie ne se limite pas à l'explication; elle est également en mesure de proposer aux individus des manières efficaces d'atteindre leurs buts[94]. Toutefois, lorsque les agents doivent se conformer à une norme sociale afin d'atteindre ces buts, un conseil rationnel, pris au pied de la lettre, serait au pire auto-destructeur, et au mieux non maximiseur. L'alternative pour un modèle rationnel serait de conseiller aux gens de ne pas paraître rationnel, c'est-à-dire de suivre les recommandations du modèle, mais en secret. Je ne vois pas comment, en théorie des jeux, on pourrait concevoir un jeu dans lequel les joueurs ne peuvent révéler aux autres que leurs choix seront rationnels. Aucune de ces critiques ne se veut dévastatrice. Les modèles rationnels du conformisme aux normes sociales n'ont pas à expliquer les comportements dans leurs moindres détails, et ils ne sont pas obligés non plus de livrer des conseils normatifs. Nous ne désirons certainement pas nous débarrasser de la théorie du choix rationnel dans le champ des normes sociales. Notre objectif a été de montrer certaines lacunes de la théorie lorsque les motivations et les pratiques reliées aux normes sociales sont à l'étude. Un bon exemple selon nous d'un modèle rationnel des normes sociales conscient de ses limites est celui de Knight et Ensminger (1998) sur les normes de mariage dans une nation africaine. Ils décrivent les pratiques en termes de marchandage pour le pouvoir socio-économique, mais ils n'essaient pas d'expliquer les rituels symboliques en soi, ce qui aurait causé la faillite de

---

[92] On retrouve cette stratégie explicative surtout dans les modèles d'attentes mutuelles. Frank (1998) maintient que la norme contre la consommation ostensible (*conspicuous consumption*) sert à prévenir l'explosion des dépenses au sein du groupe. Mais pour qu'une telle explication soit pleinement convaincante, on doit démontrer que les agents se conforment à la norme dans ce but explicite. Il semble douteux que les agents de Frank expriment du mépris envers le voisin et sa nouvelle voiture sport parce qu'ils ont calculé, même superficiellement, les conséquences à long terme d'avoir à répliquer par d'autres dépenses folles. Le conformisme pourrait être expliqué plus simplement par les valeurs égalitaires de la norme, ou encore par une manifestation de l'envie.

[93] Satz et Ferejohn (1994) préconisent une distinction formelle entre les explications micro et macro en théorie du choix rationnel.

[94] Follesdal (1982), Elster (1986a).

l'analogie du marché. A l'opposé, Brennan et Pettit (2004) s'intéressent aux pratiques, et sont ainsi conduits à une modélisation rationnelle fonctionnaliste.

Plusieurs modèles rationnels de l'échange de dons ou de cadeaux nous présentent également des explications à tendance fonctionnalistes. Ruffle (1999) nous propose un modèle fondé sur une fonction d'utilité additionnant des utilités matérielles et émotionnelles. Sa stratégie consiste à relier les émotions aux croyances anticipées. Dans la version unilatérale du jeu, le récipiendaire sera surpris s'il s'attendait à un cadeau de moindre valeur que celui qu'il a reçu, et le donateur vivra un sentiment de fierté à avoir ainsi surpris son partenaire. La rationalité des joueurs ainsi que leurs fonctions d'utilité sont présumés être de connaissance commune. Ce modèle fonctionne assez bien lorsque la jouissance du cadeau peut être séparée des motivations du donateur. Un généreux pourboire peut satisfaire la serveuse même si elle sait que la cliente a agi de manière tout à fait rationnelle car, dans ce contexte, l'utilité de l'argent n'est pas affecté par les motivations du donateur. Dans la plupart des transactions où l'on qualifie le bien donné de "cadeau", toutefois, la motivation du donateur va jouer un rôle déterminant dans la valeur du cadeau pour le récipiendaire, ce que nous nommons habituellement la "valeur sentimentale"[95]. Il nous apparaît assez clair que, si le récipiendaire sait que le donateur a agi de façon rationnelle (comme le suppose expressément le modèle), il n'y aura pas de valeur sentimentale ajoutée. Contrairement à ce que prétend Ruffle (1999: 415), des amoureux ne profiteraient pas d'un échange de cadeaux sachant que l'autre cherche à maximiser sa fierté (moins les coûts) en maximisant la surprise du conjoint. Dans ce sens précis, on pourrait parler d'une norme sociale de l'échange de cadeaux. Une telle norme se maintiendrait minimalement par l'incertitude, non pas envers quelque paramètre de la fonction d'utilité du partenaire, mais envers sa motivation, rationnelle ou non.

Notre modèle repose, entre autre, sur le principe que, dans le champ des normes sociales, *les motivations ont de l'importance*. Certains modèles rationnels, ne portant pas nécessairement sur les normes, ont tenté d'intégrer les effets émotionnels ou réputationnels de certaines motivations. Hirschleifer (1987) a cherché à modéliser la gratitude et la colère dans

[95] Voir Zelizer (1994) pour une étude sociologique des valeurs que peut prendre l'argent dans la vie quotidienne.

la réponse du second joueur à l'attitude du premier joueur dans un projet coopératif. Même s'il maintient que l'agent "can be *passionate*, in the sense of 'losing control'" (Hirschleifer 1987: 317, italiques originales), que celui-ci peut réagir différemment selon que son partenaire a agi de façon intentionnelle ou non, et qu'une réaction émotionnelle se veut "non utilitaire", en bout de ligne le modèle fait dépendre la gratitude et la colère strictement du niveau de coopération d'autrui; les motivations n'y jouent aucun rôle. Bernheim (1994) a aussi tenté d'élaborer un modèle rationnel sensible aux motivations, à travers l'allocation d'estime[96], mais les types de motivations que l'on y retrouve ne sont que des éléments d'un ensemble de fonctions d'utilités. Les agents y sont toujours motivés par la rationalité, seulement certaines variables de fonctions changent. Ces modèles, ainsi que bien d'autres, n'ont d'autre choix que de se rabattre sur les motivations rationnelles et rechercher dans les fonctions d'utilité ou l'ordre des préférences la source des variances de comportement des agents, car la théorie du choix rationnel, presque par définition, n'est pas outillée pour les motivations non rationnelles, à l'exception peut-être des comportements automatiques de type stimuli-réponse, qui peuvent être pris en charge par une théorie évolutionnaire de la rationalité, mais, comme nous l'avons vu dans l'introduction de la thèse, la théorie évolutionnaire ne se préoccupe pas des motivations.

Notre modèle du conformisme combine la rationalité, les motivations émotionnelles de Elster, et les mécanismes de reconnaissance de Bourdieu. Les avantages d'une telle approche qui nous semblent les plus significatifs sont les suivants. D'abord, notre approche n'accorde pas d'importance particulière aux bénéfices collectifs des normes, contrairement aux modèles rationnels qui, habituellement, se servent de ces bénéfices pour expliquer le conformisme[97]. Nous pouvons tout aussi bien traiter du conformisme à des normes néfastes pour la collectivité sans avoir à recourir à des bénéfices indirects et subtils, comme nous l'avons démontré avec l'équilibre hypocrite. Nous n'en négligeons pas toutefois les conséquences pour le groupe. Ils font partie du modèle comme déterminant de la stabilité de l'équilibre : les normes bénéfiques pour le groupe auront tendance à être plus stables, car, entre autres, certains agents auront

[96] "If, for example, it comes to light that some supposedly generous individual took apparently generous actions for selfish reasons, then the esteem accorded to that individual diminishes. (...) Thus status depends critically on motivations." (Bernheim 1994: 844).

[97] Des individus motivés par les valeurs peuvent consciemment chercher à maximiser le bien-être collectif, mais pas nécessairement, et du moins certainement pas en termes instrumentaux comme les coûts de transaction.

alors un incitatif supplémentaire à s'y conformer. A l'inverse, les normes néfastes auront tendance à être moins stables, car il planera toujours la possibilité qu'une masse critique d'agents se rendent compte du cercle vicieux dans lequel ils évoluent. Un autre avantage de notre modèle est son habileté à décrire les équilibres normatifs d'une manière plus étoffée que les modèles rationnels standard. Comme nous l'avons vu au chapitre II, les théories du choix social mettent l'accent sur les comportements rationnels tout en écartant les manifestations axiologiques et émotionnelles présentes dans les débats publics. A l'autre bout du spectre, les théories de la démocratie délibérative supposent que les citoyens suivent des normes d'impartialité, sans considérer la possibilité d'une exploitation rationnelle de ces normes. A notre avis, ces deux familles de théories nous apparaissent incomplètes lorsqu'il s'agit d'expliquer, ou même de prescrire, le comportement individuel. Un meilleur traitement du champ politique inclurait des motivations autant rationnelles que non rationnelles. Nous sommes convaincus qu'une théorie des normes sociales fondée sur les valeurs, les émotions et la rationalité pourrait jeter un éclairage nouveau sur biens des champs d'interaction sociale. Au chapitre suivant, nous allons tenter d'appliquer notre modèle motivationnel au champ des relations de pouvoir, plus spécifiquement au sein des théories organisationnelles. Tout comme ce fut le cas pour le conformisme aux normes sociales, nous espérons que notre analyse de l'autorité à partir de la distinction entre motivations rationnelles et non rationnelles nous permettra de saisir des phénomènes que la rationalité seule ne peut adéquatement traiter.

CHAPITRE V

LA LÉGITIMATION DE L'AUTORITÉ DANS LES THÉORIES RATIONNELLES DU POUVOIR

Pour les sciences sociales fondées sur la rationalité, saisir la complexité de la notion de pouvoir se révèle une tâche délicate. Les modèles rationalistes s'entendent pour définir le pouvoir comme une relation entre agents où A tente d'influencer le comportement de B de façon à ce que ce dernier agisse dans l'intérêt du premier. Ce genre de relation, à l'apparence simple, recèle toutefois de nombreuses subtilités concernant les intentions des agents, les raisons d'obéir, les sources matérielles et réputationnelles de pouvoir, le rôle des normes sociales, et autres. Le caractère éclectique de la notion de pouvoir ainsi que son rapport délicat avec le contenu intentionnel de part et d'autre de la relation constitue un obstacle majeur à une axiomatisation simplifiée, ce qui fait que les modèles rationnels n'ont pas vraiment d'autre choix que de recourir au "découpage" de la notion selon certaines caractéristiques des agents et de la situation, et ainsi élaborer des typologies du pouvoir plutôt qu'une seule définition formelle. Dans l'état actuel de la discipline, toutefois, ces typologies prêtent trop souvent à confusion. Outre le fait que certains auteurs emploient les mêmes termes pour désigner des phénomènes distincts, les types de pouvoir que nous rencontrons au sein d'un même modèle ont parfois tendance à se recouper.

En résumé, une autorité légitime[98] est une relation où le dominé ne se perçoit pas comme tel; il n'a pas l'impression de vivre une relation de domination. Un pouvoir brut est, pourrait-on dire, une relation "classique" où, par exemple, la menace de sanctions de la part de A incite B à obéir. Dans un tel cas, B se sait pris dans une relation de pouvoir, mais il obéit parce que l'alternative lui apparaît pire encore, alors que dans une relation légitime, l'obéissance fait partie de ses propre désirs, car il sent qu'il en va de son devoir d'agir ainsi. L'approche rationaliste du pouvoir met en jeu, d'un côté, les intérêts individuels et les

[98] Nous préférons le terme "autorité" pour désigner le pouvoir légitime car il décrit mieux son caractère acceptable. Du point de vue du dominé, être soumis à un "pouvoir" a une connotation de servitude, alors que respecter l'"autorité" montre déjà une forme de validation de la relation. Au-delà de leur utilisation dans le langage populaire, nous ne maintenons pas qu'il existe une différence fondamentale entre ces deux termes.

ressources disponibles du dominant, et de l'autre, les coûts anticipés de la désobéissance du dominé, considérant que, *ceteris paribus*, celui-ci préfère ne pas se soumettre à la volonté d'autrui et conserver son autonomie. Au chapitre précédent, nous nous sommes penchés sur la dynamique du conformisme aux normes sociales. Le problème avec l'approche rationaliste des normes sociales, c'est que les sujets savent que la maximisation individuelle d'utilité s'effectue souvent sur le dos du reste du groupe. Le conformisme aux normes sociales confère reconnaissance et réputation, mais pourvu que l'agent, par le fait même, y sacrifie son intérêt particulier au nom de l'intérêt collectif instancié par les normes. L'agent rationnel qui désire maximiser les profits du conformisme ne peut pas paraître ouvertement rationnel, il doit *faire croire* qu'il place l'intérêt du groupe au-dessus de son intérêt personnel. Nous retrouvons une dynamique semblable dans le champ du pouvoir. La relation de pouvoir la plus efficace est la relation légitime, celle qui ne se présente pas ouvertement comme un exercice de pouvoir direct. Même si l'on peut expliquer toute relation de pouvoir par les intérêts, le contrôle de ressources, et les calculs coûts/bénéfices[99], il faut tenir compte qu'en tant que *motivation* du dominant, ce type de pouvoir est la plupart du temps inacceptable. Le dominant sera donc amené à reformuler la relation, en ayant recours aux élément de la structure comme les lois, les règles et les normes. Une relation de pouvoir se révèle beaucoup plus acceptable lorsque le dominant, au lieu de dire "j'exige que...", est en mesure de dire "le règlement exige que...", "de par mon titre, je suis en droit d'exiger que...", ou encore, "le groupe exige que...". Tout comme pour les normes sociales, une explication rationnelle des relations de pouvoir est toujours possible, mais elle demeurera en quelque sorte fonctionnellement rationnelle, tant et aussi longtemps que les sujets ne *veulent* pas qu'une telle explication s'applique à eux, et qu'ils ajustent leurs comportements en conséquence. C'est ce type de comportement qu'une approche exclusivement rationnelle ne peut saisir.

Les théories rationnelles du pouvoir mettent surtout l'accent sur l'efficacité de l'autorité légitime. Si le dominé considère la relation comme étant légitime, le dominant n'aura pas besoin de recourir aux sanctions et à la surveillance pour s'assurer que l'autre se comporte selon ses

[99] C'est la conception de Coleman (1986) du pouvoir, en accord avec sa position sur les normes sociales. Dans un exercice légitime du pouvoir, le dominant fait profiter son groupe de ses ressources, alors que dans une relation illégitime, le dominant "vole" le groupe en s'accaparant des ressources, soit pour lui-même ou pour les redistribuer à l'extérieur du groupe.

volontés. Nous proposons d'aller plus loin en affirmant que la légitimation d'une relation de pouvoir n'est pas seulement une question d'efficacité, mais que bien souvent, les normes sociales en vigueur au sein du groupe rendent cette légitimation nécessaire. L'exercice brut de pouvoir est rarement considéré comme une relation interpersonnelle acceptable, sauf dans certains champs bien circonscrits et socialement justifiés, comme la famille ou la prison. Lorsque le pouvoir du dominant n'est pas absolu, c'est-à-dire que le dominé, ou une tierce partie, est en mesure de réagir avec succès contre la tentative de domination, la légitimation devient une *condition* de l'exercice du pouvoir, et non seulement une stratégie plus rentable. La structure complexe de pouvoirs et de contre-pouvoirs s'exprime à travers des normes précisant les limites de l'acceptable. De là toute l'importance de la distinction entre pouvoir nu et pouvoir légitimé.

Dans cet article, nous proposons une nouvelle manière de conceptualiser l'autorité légitime, tout en demeurant, autant que possible, près des typologies rationnelles. Pour ce faire, nous survolerons d'abord les approches de quelques théories rationnelles du pouvoir qui ont marqué la discipline. Nous nous intéresserons de plus à une théorie du pouvoir légitime se situant en-dehors du paradigme rationnel, celle du pouvoir symbolique de P. Bourdieu. A partir d'une critique de ces modèles, nous tenterons d'échafauder une alternative viable. Ensuite, nous examinerons l'application du concept de légitimité du pouvoir dans un champ concret. Nous avons choisi d'étudier les relations de pouvoir en entreprise, telles que modélisées par les théories rationnelles des organisations et du management. L'intérêt premier du milieu de travail réside dans sa codification souvent stricte et sans ambiguïtés des normes régissant les relations entre agents, ici présentés sous forme d'organigrammes hiérarchiques et de règles écrites. En d'autres termes, et nous reviendrons fréquemment sur ce point, cette structure formelle de légitimation se veut beaucoup plus visible au sein de l'entreprise que, par exemple, dans les relations informelles en société où les normes sociales sont présumées connues de tous, mais jamais codifiées. En conclusion, nous reviendrons sur le concept d'autorité légitime à la lumière de ce que la théorie des organisations nous aura appris, en tenant compte des particularités d'un tel champ d'application.

5.1 L'autorité légitime

Les théories relationnelles

Fondamentalement, les théories du pouvoir d'inspiration rationnelle traitent de relations de pouvoir entre au moins deux acteurs, où les préférences et les actions du dominant peuvent agir sur les préférences et les actions du dominé de manière à ce que ce dernier soit amené à servir les intérêts du premier[100]. On distingue de manière générale l'école élitiste, ou néo-élitiste, qui situe le pouvoir dans une classe d'individus caractérisée par des facteurs socio-économiques et institutionnels, et l'école pluraliste qui soutient que tous les agents peuvent exercer un certain pouvoir. Comme toutes deux ont recours à peu près aux mêmes concepts de base concernant le pouvoir, nous ne jugeons pas utile de conserver cette distinction[101]. Nous les nommerons, suivant Goetschy (1981), "théories relationnelles", car elles ont en commun de penser le pouvoir comme une relation entre agents. La forme de pouvoir la plus directe, celle sur laquelle à peu près tous les théoriciens s'entendent pour y accoler l'étiquette "pouvoir", implique l'usage de sanctions de la part de A dans le but de contraindre B à exécuter une action qui bénéficie à A et qu'il n'aurait pas faite autrement. Dans la forme "pure", le dominé demeure pleinement conscient de la nature de la relation; la domination est vécue *comme* une relation de pouvoir. Ce type de relation est la plus difficile à maintenir. En plus des coûts de maintien de l'appareil de sanctions, il faut considérer les effets de cet antagonisme sur le rendement du dominé, qui verra à en faire le moins possible et à se sortir de la relation au plus vite. Une relation de pouvoir dans laquelle le dominé ne se perçoit pas comme agissant contre son gré se révélera beaucoup plus efficace, car les sanctions s'avéreront inutiles et la participation de l'agent sera en quelque sorte voulue ou préférée. C'est ce que nous entendons en gros par "autorité légitime", et nous nommerons "légitimation", le passage d'une relation brute de pouvoir à un pouvoir légitime. Il ne s'agit pas ici d'une appréciation morale de la relation par l'observateur externe, mais bien de la perception du dominé de sa propre situation.

---

[100] Quelques définitions formelles : "Influence is a relation among actors such that the wants, desires, preferences, or intentions of one or more actors affect the actions, or predispositions to act, of one or more other actors" (Dahl 1991: 32). "A power relation, actual or potential, is an actual or potential causal relation between the preferences of an actor regarding an outcome and the outcome itself. (...) [T]he *outcome* must be a variable indicating the state of another social entity - the behavior, beliefs, attitudes, or policies of a second actor" (Nagel 1975: 29, italiques originales).

[101] Voir Cox, Furlong et Page (1985: 80-121) pour une comparaison détaillée de ces deux courants.

Les théories relationnelles du pouvoir que nous aborderons ici cherchent avant tout à constituer des typologies du pouvoir. Parmi l'ensemble des types de pouvoir que nous considérons comme légitimes, Bachrach et Baratz (1970) établissent une distinction entre l'*influence*, émanant du respect du dominé pour les attributs ou les titres du dominant, et l'*autorité*, résultant d'une délibération raisonnable entre les agents. Dans une relation d'autorité, le dominé reconnaît le caractère raisonnable de l'ordre à la lumière de ses propres valeurs. Ces deux relations sont qualifiées de "rationnelles", au sens où elles impliquent un choix de la part du dominé d'obéir ou non. La *manipulation* est un type de pouvoir non rationnel où le dominant tente d'obtenir l'assentiment de sa victime par des arguments mensongers. Celle-ci n'a pas le choix d'obéir, puisqu'en principe, elle ne sait même pas qu'elle se fait manipuler[102]. Nous retrouvons également cette distinction entre délibération raisonnable et argumentation mensongère chez Dahl (1991) et Wrong (1988), sous le vocable de *persuasion* et de *manipulation*, respectivement. Wrong distingue en plus la notion d'autorité, qui signifie pour lui le respect de la source de l'ordre, plutôt que son contenu sémantique[103]. Il établit cinq types d'autorité; les deux premiers (autorité coercitive et autorité incitative) concernent l'usage de sanctions et ne respectent pas nos critères de légitimité. L'*autorité légitime* constitue la reconnaissance, de la part du dominé, du droit du dominant de commander, et de son propre devoir d'obéissance, ce qui suppose le partage de certaines normes sociales. Cette reconnaissance porte uniquement sur les rôles sociaux des agents, et non sur les arguments du dominant. Les deux derniers types sont l'*autorité compétente*, soit la reconnaissance de l'expertise du dominant, et l'*autorité personnelle*, soit le respect de ses qualités propres. Pour prendre un dernier exemple, Raven (1965) distingue entre l'influence informationnelle (qui ressemble à la persuasion rationnelle), le pouvoir de l'expert, la volonté d'émulation d'autrui, et l'influence légitime, qui correspond à l'autorité légitime de Wrong.

---

[102] Nous avons pris certaines libertés avec la typologie de Bachrach et Baratz, car en réalité ils distinguent le concept de pouvoir de trois autres concepts, soit la force, l'influence et l'autorité; la manipulation faisant partie de la force. Ils réfèrent à cette famille comme "le pouvoir et les concepts apparentés", ce qui peut porter à confusion, car le lien de parenté n'est jamais clairement explicité. Considérer les trois types mentionnés plus haut comme types de pouvoir ne correspond pas à la typologie originale, mais elle nous permet la comparaison avec les auteurs qui suivront.

[103] Il faut être prudent ici, car l'usage du terme "autorité" chez Bachrach et Baratz, et chez Wrong, n'est pas du tout le même.

Ces typologies nous semblent toutefois trop fines pour tenir la route. Bien que la notion même de pouvoir soit floue, et que par conséquent toute typologie comportera ses exceptions et ses zones grises, il nous semble que les frontières proposées ici sont particulièrement poreuses. Un exemple classique est celui du patient qui obéit aux ordres de son médecin. Qu'est-ce qui le motive ainsi ? Nous pourrions affirmer qu'il est influencé par la persuasion rationnelle (peut-être mêlée d'un peu de manipulation...), soutenue par son autorité compétente, elle-même supportée par son titre officiel (autorité légitime), et peut-être son propre charisme y joue-t-il un rôle... A noter qu'il ne s'agit pas seulement de montrer par cet exemple que des formes multiples de pouvoir peuvent opérer simultanément; il nous indique que ces formes s'influencent entre elles. Notamment, pour reprendre la distinction de Wrong, le contenu sémantique de l'ordre est presque toujours interprété à la lumière du statut social et des attributs personnels de celui qui l'énonce. Galbraith (1983) emploie pour sa part une typologie plus minimaliste. Il distingue entre le recours aux sanctions négatives et positives, et le *pouvoir conditionné*, lorsque "le consentement à l'autorité, la soumission à la volonté d'autrui, devient la plus haute préférence de celui qui se soumet" (Galbraith 1983: 24). La soumission devient alors "le produit du propre sens moral ou social de l'individu" (Galbraith 1983: 35). Toute relation de pouvoir conditionné se situe quelque part entre deux pôles, le *conditionnement explicite*, qui englobe le recours à des méthodes telles la persuasion rationnelle ou non (au sens évoqué plus haut), l'éducation, le marketing, etc., et le *conditionnement implicite*, "dicté par la culture elle-même; la soumission étant considérée comme normale, appropriée, ou traditionnellement correcte" (Galbraith 1983: 24). L'idée d'un continuum est intéressante, car elle suppose, en quelque sorte, que le recours aux valeurs du dominé par la persuasion doive passer par la culture du dominé. Galbraith souligne également un aspect crucial de la légitimité du pouvoir : les valeurs communes d'une société peuvent interdire certaines formes de conditionnement explicites, forçant le dominant à respecter ces valeurs, donc à pencher vers un conditionnement implicite. Cet aspect est repris d'une façon beaucoup plus fouillée par la théorie du pouvoir symbolique de Bourdieu.

Le pouvoir symbolique

Il nous faut présenter, avant de procéder à l'examen du pouvoir symbolique, la mécanique générale proposée par Bourdieu des comportements individuels en société, qui

repose sur la distinction entre la pratique et sa représentation. La *connaissance pratique* est une connaissance pré-réflexive, non formalisée par l'agent. Les motivations à l'action qu'elle engendre sont de nature émotive : l'agent ne sait pas tout à fait ce qui le pousse à agir d'une certaine façon, mais il ressent des émotions positives à agir ainsi, et inversement, des émotions négatives lorsqu'il va à l'encontre de "ce qu'il faut faire" (Bourdieu 2003: 265-67). La *logique pratique* est une forme d'"économie" mettant en scène des agents mus par la connaissance pratique de leur position sociale. Elle se distingue de la rationalité en ce qu'elle n'a "ni la *rigueur* ni la *constance* qui caractérisent la logique logique, capable de *déduire* l'action rationnelle des principes explicites et explicitement contrôlés et systématisés d'une axiomatique (...)." (Bourdieu 1980: 174, italiques originales). La *représentation symbolique* de la pratique consiste à exprimer, publiquement, une certaine interprétation de la pratique. Une description totale de la pratique s'avère impossible, car même le sujet n'y a pas complètement accès. Il y aura donc, en tout temps, plusieurs représentations possibles pour une même pratique; et, de là, la possibilité de luttes politiques pour la représentation la plus juste (Bourdieu 2001: 193-94). La représentation symbolique s'apparente à un type de norme sociale, qui prendrait la forme d'expression de règles mettant les valeurs communes des agents en jeu. Ces agents découvrent ainsi une manière plus formelle d'agir au nom de ces valeurs. Cette formalisation sert également de point focal, de lieu de rencontre entre personnes partageant les mêmes valeurs, créant ainsi le groupe. Pour Bourdieu, le glissement de l'agent vers un respect immédiat et non rationnel de la norme, passage de l'intérêt au désintéressement facilité par la *self-deception*, est un acte de *méconnaissance*, "se reconnaître à tort dans une forme particulière de représentation et d'explicitation publique de la *doxa*" (Bourdieu 2003: 267).

Le pouvoir se fonde chez Bourdieu sur la notion de "capital". Les agents en situation d'interaction possèdent chacun un certain capital, entendu comme une capacité de mobilisation de ressources dans un champ particulier. Ce capital prend différentes formes selon le champ, on parle alors de capital matériel, religieux, littéraire, scolaire, etc. Les interactions sociales s'effectuent selon la logique économique, en fonction des rapports de force conférés par la détention de capitaux dans les champs où ils ont de la valeur. Toutefois, dans les champs autres que purement économiques, des normes sociales viennent interdire l'usage *manifeste* des rapports de force au détriment du respect du groupe. Il devient alors nécessaire de redécrire ce

genre de transactions en termes convenables. Un capital, de quelque forme que ce soit, qui est "méconnu et reconnu" comme légitime, donc conforme aux normes sociales en vigueur, devient un *capital symbolique* (Bourdieu 1980: 209-10). Le capital symbolique est le "produit de la transfiguration d'un rapport de forces en rapport de sens", il "n'est pas une espèce particulière de capital mais ce que devient toute espèce de capital lorsqu'elle est méconnue en tant que capital, c'est à dire en tant que force, pouvoir ou capacité d'exploitation (actuelle ou potentielle), donc reconnue comme légitime" (Bourdieu 2003: 347). Synonyme d'honneur et de réputation, il constitue ce "profit" de reconnaissance du groupe lorsque l'agent se plie, ou semble se plier, aux exigences normatives du groupe[104]. Le groupe est le lieu où le besoin d'estime d'autrui est comblé : l'appartenance au groupe, s'exprimant dans l'obéissance (réelle ou apparente) aux préceptes du groupe, est récompensée par la reconnaissance[105].

L'expression de l'intérêt du dominant en termes de valeurs et d'intérêts communs aux membres du groupe, et la reconnaissance par ces derniers de la légitimité de la redescription, c'est ce que Bourdieu nomme la *violence symbolique*, ou la *soumission doxique* (Bourdieu 2003: 245-48; 1994: 126-27). Il ne s'agit pas d'une domination justifiée en termes raisonnables, permettant aux sujets de délibérer et d'accepter librement une certaine aliénation de leur volonté au nom d'un but collectif concret (gain de productivité, défense contre l'ennemi commun, etc.). La violence symbolique suscite directement des valeurs fortement intériorisées ("durablement inscrites dans les corps") qui appellent à la soumission immédiate, et qui se présentent comme "la seule chose à faire". Le capital, comme capacité de mobilisation, est

---

[104] Les propriétés psychologiques du dominant font également l'objet de reconnaissance : "La reconnaissance est le résultat d'un processus par lequel les acteurs sociaux accordent à des individus ou des groupes des qualités spécifiques dont ils se croient eux-mêmes dépourvus, actualisant du même coup leur propre soumission à la domination" (Voirol 2004: 413). Bourdieu nomme ce principe "idolâtrie politique", "(...) la valeur qui est dans le personnage politique, ce produit de la tête de l'homme, apparaît comme une mystérieuse propriété objective de la personne, un charme, un charisme; le ministerium apparaît comme mysterium" (Bourdieu 2001: 261). Ceux qui s'identifient au dirigeant aiment croire que c'est son caractère, et non ses moyens financiers ou l'organisation autour de lui, qui l'a propulsé au sommet (Galbraith 1983: 41-44).

[105] C'est aussi pour Bourdieu une question d'identité : "L'effet d'oracle (...) est ce qui permet au porte-parole autorisé de s'autoriser du groupe qui l'autorise pour exercer une contrainte reconnue, une violence symbolique, sur chacun des membres isolés du groupe. Si je suis (...) le groupe fait homme, et si ce groupe est le groupe dont vous faites partie, qui vous définit, qui vous donne une identité, qui fait que vous êtes vraiment un professeur, vraiment un protestant, vraiment un catholique, etc., il n'y a plus vraiment qu'à obéir." (Bourdieu 2001: 270).

source de pouvoir. Le capital symbolique étant la légitimation d'une forme particulière de capital, le *pouvoir symbolique* est la légitimation de son utilisation, c'est-à-dire de l'exercice du pouvoir que ce capital particulier confère. Le pouvoir symbolique est donc relié au pouvoir politique, économique, etc., mais il ne s'y réduit pas, car dans les champs sociaux où l'exercice brut du pouvoir n'est pas acceptable, l'euphémisation des relations de pouvoir devient incontournable (Swartz 1997: 89). L'exercice du pouvoir symbolique consiste à amener autrui à lui obéir, non pas à cause de sanctions ou d'incitatifs directs, mais en faisant correspondre l'acte demandé aux valeurs des agents. Ce type de redescription est rendu possible par la notion de "connaissance pratique", qui stipule qu'il n'existe pas de descriptions précises de nos valeurs. Bourdieu relève deux facteurs d'efficacité du pouvoir symbolique : la réputation du dominant et la véracité de la représentation symbolique qu'il propose (Bourdieu 1987: 163-64), ce qui nous donne deux arènes de luttes symboliques possibles, pour la réputation et pour la représentation authentique du groupe dans un contexte où elle ne sera jamais qu'approximative.

L'autorité légitime revisitée

La légitimation du pouvoir est le passage de l'autorité directe et ouvertement intentionnelle à une forme "naturelle" d'autorité, qui se présente au dominé comme "la chose à faire". Cette stratégie fait entrer en scène la structure sociale partagée par les dominants et les dominés. Lukes (1977) établit une distinction entre la *structure* et le *pouvoir*. La structure influence directement le comportement, sans passer par les raisons du dominé. Elle peut prendre la forme de contraintes physiques, économiques, etc., ou encore se présenter comme un ensemble de règles et de normes (Lukes 1977: 12-14). L'exercice du pouvoir s'effectue en deçà de ces contraintes structurelles, au niveau des contraintes résiduelles que le dominé est en mesure d'évaluer comme des raisons de se soumettre. Lorsque le dominé est aux prises avec ces contraintes dites "rationnelles", au sens de "fondées en raison", il peut toujours choisir la soumission, alors que cela ne peut être le cas au niveau des contraintes structurelles. Les sanctions, positives ou négatives, font partie des contraintes rationnelles, car elles constituent le "prix" de l'action, encouragée ou proscrite, qui entrera dans le calcul rationnel de l'agent. Le concept de structure demeure toutefois subjectif : ce qui est contrainte structurelle pour un peut être contrainte rationnelle pour l'autre.

Nous proposons de fonder la notion d'autorité légitime sur le degré d'influence structurelle perçu par le dominé. Plutôt que de séparer nettement "pouvoir" et "structure", nous envisageons l'exercice d'un pouvoir qui tire sa force d'une exploitation de la structure[106]. Considérant la nature relationnelle du pouvoir, si la "force" dominante s'avère être perçue comme étant de nature structurelle[107], on ne retrouve pas d'agent dominant (au sens d'une personne), donc pas de relation de pouvoir. La légitimation de la structure peut prendre diverses formes. Dans la structure *formelle*, le dominé perçoit une règle ou une loi comme légitime lorsqu'elle ne lui apparaît pas comme un instrument aux mains d'un ou de plusieurs agents servant à l'exercice d'un pouvoir[108]. Une autorité médiée par la structure *informelle*, soit les normes sociales, se veut légitime lorsque la norme fait appel aux valeurs des agents, car celui qui agit au nom de ses propres valeurs ne s'imagine certainement pas être la victime d'une domination. L'acte de légitimation consiste à dépersonnaliser la relation de pouvoir en la fixant en quelque sorte sur un élément de la structure du champ dans lequel évoluent dominants et dominés, en faisant appel, soit aux règles formelle, soit aux normes informelles[109].

Rappelons la définition du pouvoir symbolique : A obtient l'assentiment de B en euphémisant la relation de pouvoir afin qu'elle ait l'apparence de correspondre aux valeurs et aux normes sociales chères à B. Celui-ci, en retour, méconnaît la relation, c'est à dire qu'il en reconnaît la légitimité tout en ignorant, ou en ne voulant pas savoir, sa véritable nature. Selon les typologies des théories relationnelles, il s'agirait d'une forme de manipulation ou de fausse autorité. Ce n'est pas une forme d'autorité légitime comme l'entend Wrong, car, pour lui, c'est

[106] Voir Chazel (1983) pour une discussion similaire sur la complémentarité entre la domination, de nature structurelle, et le pouvoir, de nature relationnelle. Il semble toutefois réserver la domination à une distribution inégale de ressources, sans considérer l'aspect interne de la structure du modèle de Lukes, dont il est d'ailleurs assez critique.

[107] Notre emploi du terme "structure" n'implique pas de notre part une adhésion à un cadre d'analyse structuraliste. Le terme désigne simplement un type de contrainte à l'action qui n'est pas perçu comme provenant directement d'un autre agent.

[108] Par exemple, on dit qu'une loi est "légitime" lorsqu'elle sert à maintenir l'ordre (un bien public), et "illégitime" lorsqu'elle sert visiblement les intérêts des législateurs au détriment du reste de la société.

[109] Si le dominant peut en appeler aux *intérêts* du dominé plutôt qu'à ses valeurs, en lui faisant croire que l'obéissance est dans son meilleur intérêt, alors la relation devient une instance de pouvoir incitatif. Ce type de pouvoir se veut certainement plus acceptable qu'un pouvoir coercitif, mais il représente néanmoins une relation de pouvoir perçue comme telle par le dominé, tandis qu'avec l'autorité légitime, la relation, en bonne partie du moins, ne se vit pas comme une domination.

la *source*, celui qui ordonne, qui se voit légitimée et non pas le *contenu*, ce qui serait un cas de persuasion (Wrong 1988: 49). Or, chez Bourdieu, source et contenu se confondent; le discours étant en grande partie un produit de la position sociale du locuteur. C'est l'euphémisation de la relation qui est opérante. Les symboles, les discours et la réputation constituent les ressources du dominant dans l'exercice du pouvoir symbolique, tout comme le contrôle de ressources matérielles permet l'exercice d'un pouvoir incitatif. Pour Wrong, les raisons pour lesquelles le pouvoir nu cherche à se transformer en autorité légitime sont, premièrement, que la réaction du dominé devient plus prévisible et deuxièmement, que le dominant n'a pas besoin de maintenir un coûteux système de sanctions (Wrong 1988: 52). Le recours au pouvoir symbolique n'est pas directement relié à de semblables avantages. L'euphémisation de la relation de pouvoir est rendue *nécessaire* par la présence d'une norme sociale interdisant l'exercice d'un pouvoir coercitif ou incitatif. Il ne s'agit donc pas d'un choix de stratégie de domination. Le prêtre qui veut attirer les croyants à la messe ne peut pas s'y prendre par la contrainte physique, ni par le paiement en espèces; sa congrégation ferait long feu. Le champ social du rituel religieux contient une norme interdisant ce genre de relations. Dans le pouvoir symbolique, le capital matériel est inutilisable car il ne constitue pas un "paiement" acceptable pour le dominé, à moins qu'il ne soit justifié au nom de ses propres valeurs ou des normes sociales existantes, autrement dit, qu'il soit converti en capital symbolique.

Dans notre conception de l'autorité légitime, l'euphémisation, bien que jouant un rôle important, n'est pas nécessaire. Selon la situation, le capital matériel peut conserver sa valeur; le dominé peut à la fois accepter la rémunération matérielle et agir au nom de ses valeurs, ou du moins rationaliser la relation afin de rendre le paiement compatible avec ses valeurs. Nous ne considérons pas la légitimation du pouvoir comme toujours nécessaire. Elle peut se présenter comme une stratégie parmi d'autres pour le dominant; une stratégie qui, toutefois, s'avérera bien souvent plus efficace que la coercition directe. La force d'une autorité légitime informelle dépend des valeurs qu'elle met en jeu. Si des valeurs moins fortement intériorisées sont invoquées de façon crédible par le dominant[110], le sujet peut choisir de ne pas obéir, mais

[110] Si le sujet ne considère pas crédible l'appel aux valeurs du dominant, si en d'autres termes il n'y voit qu'une rhétorique creuse, ses valeurs ne seront pas engagées et ne contraindront pas sa réaction. Elles pourraient toutefois engendrer de la colère face à leur usurpation opportuniste.

il devra alors fournir, à lui-même et aux autres, une bonne justification[111]. Le défi d'honneur provenant d'un supérieur ne peut être refusé que d'une manière honorable pour quiconque possédant une valeur modérée d'honneur, et ne sera jamais refusé par un sujet ayant fortement incorporé l'honneur. Pour pousser l'exemple plus loin, celui qui ne joue pas le jeu de l'honneur peut platement refuser, sans besoin de justification.

Dans les typologies relationnelles que nous avons retenues, l'autorité légitime est certes reconnue, mais jamais méconnue. Le dominé y reconnaît soit le rôle social du dominant dans une structure hiérarchique (Raven, Wrong), ou la validité de l'argumentation raisonnable (Bachrach et Baratz), mais la relation de pouvoir demeure explicite. Galbraith relève selon nous une caractéristique capitale de la légitimation de l'autorité : moins la relation de pouvoir est visible, plus elle se voit légitimée. Dans le continuum qu'il propose, le conditionnement explicite n'est pas méconnu, comme dans la publicité où la tentative de manipulation demeure visible, alors qu'à l'autre bout, le conditionnement implicite, "dicté par la culture", est entièrement méconnu; il soutient que le conditionnement implicite est la plus efficace des relations de pouvoir. Galbraith ne nous offre pas toutefois une définition claire du conditionnement. Nous avons vu chez Bourdieu une élaboration plus complète de la relation de pouvoir méconnue. Toutefois, il s'intéresse presque exclusivement à la légitimation par la structure informelle des normes et des valeurs. Notre conception du pouvoir légitime s'accorde avec l'idée de continuum de Galbraith, dont l'extrémité implicite est éclairée par Bourdieu, en proposant en plus une distinction entre légitimation formelle et informelle car, comme nous le verrons plus clairement avec les théories du pouvoir en organisation, la dynamique du pouvoir n'est pas la même dans les deux cas.

Nous allons maintenant passer à l'étude de l'autorité légitime dans un contexte plus concret, l'entreprise. Ce type d'organisation a la particularité de posséder un ensemble de règles formelles d'une large portée et une structure hiérarchique bien définie, ce qui nous permettra de voir comment nous pourrions conceptualiser l'aspect formel de la légitimation. Nous allons

[111] Bourdieu semble abonder dans ce sens : "La connaissance que procure l'incorporation de la nécessité du monde social (...) n'exclut pas - comment peut-on croire le contraire ? - des formes de résistances (...). Mais elle reste exposée au ***détournement symbolique***, contrainte qu'elle est de s'en remettre à des porte-parole, responsables exclusifs de cette sorte de saut ontologique de la *praxis* au *logos* (...)." (Bourdieu 2003: 267, italiques originales).

tenter de voir quelle forme prend la légitimation du pouvoir dans un tel contexte, et quelles conclusions nous pourrions en tirer en vue de l'élaboration du concept d'autorité légitime.

5.2 La légitimation du pouvoir en organisation

Les théoriciens des organisations s'intéressent depuis longtemps à la question de la légitimation de l'autorité. De nombreuses écoles de pensées ont émergé à ce sujet. Au lieu de tenter de cartographier ce vaste champ, nous avons choisi de nous en tenir à quelques classiques du *management* se situant près de la théorie du choix rationnel et des typologies du pouvoir que nous venons d'examiner. Après un survol de ces modèles de pouvoir, nous approfondirons l'aspect du pouvoir en organisation qui se révèle, selon nous, le plus intéressant, soit la présence d'une structure formelle forte régissant les relations entre les membres.

Modèles organisationnels d'autorité légitime

Dans l'entreprise, une chaîne de commandement considérée comme légitime par ses membres, particulièrement ceux se situant à la fin, gagne en efficacité en vue de la maximisation des buts collectifs. L'entreprise économise ainsi sur les coûts de surveillance et de sanctions. C'est dans cet objectif que Simon (1983) propose son propre modèle. Dans cette conception particulière de la rationalité, l'agent décide à partir de deux types de prémisses : les *valeurs*, portant sur les fins, et les *faits*, portant sur les moyens pour y arriver (Simon 1983: 12-14). Outre de son profil psychologique, les valeurs de l'agent peuvent provenir de son sentiment d'identification à l'organisation, ou de l'extérieur, imposées par une autorité supérieure. Il s'agit respectivement d'un phénomène de *loyauté*, et d'un phénomène d'*autorité*. Afin de maximiser l'efficacité de l'organisation, soit d'assurer la meilleure coordination des membres dans un univers complexe, il est nécessaire de rendre l'agent le plus "rationnel" possible; ce qui signifie pour Simon l'élimination autant que possible de l'aspect "psychologique" de l'agent pour ne conserver que la loyauté et l'obéissance à l'autorité[112]. En bout de ligne, l'agent ne doit décider qu'en fonction des besoins de l'organisation : "Au niveau

[112] "L'individu rationnel est, et doit être, un individu 'organisé et institutionnalisé'. Si les limites sévères qu'impose la psychologie à la décision sont supprimées, l'individu qui prend des décisions doit rester soumis à l'influence du groupe organisé dont il fait partie. Ses décisions doivent non seulement être le fruit de ses processus mentaux mais également refléter les préoccupations plus larges qu'il revient au groupe organisé de maîtriser" (Simon 1983: 92).

psychologique, le simple fait de comprendre que tel ou tel comportement s'intègre dans un plan doit constituer un stimulus suffisant pour que l'individu s'y conforme" (Simon 1983: 111).

La différence fondamentale entre la loyauté et l'autorité est que dans le premier cas, c'est l'agent qui prend l'initiative d'ajuster ses valeurs alors que dans le second, c'est un supérieur qui s'en charge. La loyauté permet une plus grande efficacité décisionnelle, particulièrement dans les situations de crise où les routines ne fonctionnent plus et où il s'agit d'être créatif. Mais elle ne règle pas les problèmes de la coordination de la prise de décision, et aussi, les agents ayant naturellement tendance à s'identifier à leur sous-groupe et à en exagérer l'importance, elle peut amplifier les conflits pour l'attribution des ressources de l'entreprise. Pour pallier à ces problèmes, l'autorité s'avère nécessaire. Dans une relation d'autorité, l'agent "accepte de plier son comportement aux décisions d'un supérieur, sans examiner indépendamment les mérites de ces décisions" (Simon 1983: 12). Il ne s'agit pas d'un pouvoir coercitif où le supérieur obtient la collaboration du subordonné par la menace de sanctions, mais bien d'un "critère de décision" du subordonné, qui détermine une "zone de consentement" à l'intérieur de laquelle il accepte de se soumettre sans conditions. C'est une relation qui peut se définir en termes "purement objectifs et behavioristes" (Simon 1983: 112). Bien que la zone de consentement s'établisse à partir du profil psychologique de l'agent et des sanctions, positives et négatives, provenant du supérieur, une fois la zone constituée, ces facteurs ne jouent plus de rôle significatif dans le lien d'autorité, en autant que cette zone soit respectée (Simon 1983: 133-35).

Simon a voulu ramener les facteurs socio-psychologiques au sein de la science des organisations, mais ce n'était au fond que comme simples contraintes, au même titre que les contraintes physiques, économiques et technologiques pesant sur l'organisation (Simon 1983: 134). Les décisions teintées de "psychologie" ne peuvent pas être efficaces, il faut que l'agent s'en départisse en s'intégrant à l'organisation et en acceptant l'autorité de sa hiérarchie. Ce que Simon appelle les "valeurs sociales" sont en fait les valeurs suprêmes de l'organisation; il n'y a rien au dehors qui compte vraiment (Simon 1983: 178-79). Prises ensemble, la loyauté et l'autorité ressemblent à ce que nous entendons par autorité légitime, soit un pouvoir sans besoin de sanctions dans lequel le droit de commander est reconnu et l'obéissance perçue

comme un devoir. Simon a toutefois une vision très mécaniste de l'autorité légitime, croyant qu'il ne s'agit que d'accorder les valeurs des agents aux valeurs organisationnelles pour que l'entreprise accroisse son efficacité.

Bien que Simon ait inspiré toute une génération de théoriciens des organisations[113], ceux-ci ont généralement préféré dévier de son principe strict d'efficacité pour se pencher plus sérieusement sur ce qu'il entendait par "psychologie". Ils ont notamment relevé que les membres de l'organisation ont leurs propres valeurs et leurs propres buts qui peuvent entrer en conflit avec ceux de l'organisation. Galbraith (1968) établit, à l'intérieur de ce qu'il appelle la "technostructure", sa propre "théorie générale de la motivation". Cette théorie propose quatre types de motivations à obéir aux directives de ses supérieurs. Outre les deux formes de sanctions directes que sont la *contrainte* et la *rémunération pécuniaire*, nous retrouvons l'*identification* et l'*adaptation*, qui sont plus proches de l'autorité légitime. L'identification correspond à la loyauté de Simon; l'agent "juge le but assigné au groupe supérieur aux buts qu'il poursuivait lui-même antérieurement et il le fait sien" (Galbraith 1968: 140). Avec l'adaptation, Galbraith pousse la psychologie organisationnelle un peu plus loin : c'est la volonté de se soumette aux objectifs de l'organisation dans l'espoir de les influencer un jour. Cette motivation peut surgir même si l'agent *n'est pas d'accord* avec les buts de l'organisation (Galbraith 1968: 140-41). Pour Galbraith, et contrairement à Simon, les valeurs individuelles ne sont pas des obstacles à circonscrire au nom de l'efficacité organisationnelle, mais plutôt des réalités sur lesquelles il faut compter, et qui seront à la source des luttes politiques internes.

Le modèle des "cercles concentriques" de Galbraith permet de voir que les différents groupes au sein de l'organisation ne sont pas motivés par les mêmes raisons (Galbraith 1968: 158-61). Les actionnaires, situés à la périphérie, sont motivés par le gain et n'affichent que très peu de loyauté envers l'entreprise. Les cadres et la direction, au centre du cercle, sont plutôt motivés par l'identification et l'adaptation. Les motivations des ouvriers s'avèrent intéressantes : le salaire demeure le facteur premier, mais ceux-ci sont prêts à s'identifier aux buts de l'entreprise lorsqu'ils touchent un salaire raisonnable et que "la firme paraît de façon plausible poursuivre quelque autre but que de gagner le plus d'argent possible pour les

[113] La première édition de son "Administrative Behavior" est parue en 1945.

actionnaires ou la direction" (Galbraith 1968: 160). C'est le "paradoxe de la motivation économique", où "la maximisation du profit, considérée comme un but, exige que l'individu membre de la technostructure subordonne son intérêt financier personnel à celui d'actionnaires lointains et inconnus" (Galbraith 1968: 179). Ainsi, pour que l'identification puisse fonctionner, la direction doit faire appel aux valeurs des employés en redécrivant les buts véritables de l'entreprise. Il s'ensuit un double langage, car il faut continuer à persuader les actionnaires que l'entreprise recherche les profits[114].

La thèse principale de Galbraith est que, dans l'organisation moderne, la principale source de pouvoir ne se situe plus au sommet de la hiérarchie, mais bien au niveau des ouvriers spécialisés et des cadres moyens, car ce sont eux qui détiennent l'expertise et l'information nécessaires au fonctionnement de l'entreprise. Ce sont eux qui exigent sécurité et croissance avant les profits. Les phénomènes d'identification et d'adaptation font en sorte que les valeurs de l'organisation ont tendance à s'accorder avec les valeurs de la société et de chacun des membres individuellement; c'est le "principe de cohérence" (Galbraith 1968: 167-71). Par l'identification, l'agent intériorise les buts de l'organisation autant qu'il force celle-ci à respecter ses propres valeurs, et il propagera subséquemment ces valeurs organisationnelles dans le reste de la société, tandis que par l'adaptation, l'agent peut occuper un poste suffisamment élevé pour pouvoir y implanter ses propre valeurs, qu'il aura fort probablement puisé dans la société en général. Il y a donc osmose entre les valeurs propres à l'organisation, les valeurs professionnelles des employés, et les valeurs de la société dans laquelle évolue l'organisation.

Galbraith relève deux phénomènes sociaux allant à l'encontre de l'explication de l'autorité légitime en termes d'efficacité : les agents se présentent dans une organisation avec leurs propres valeurs et leurs propres buts qui ne sont pas nécessairement compatibles avec ceux de l'organisation, et l'élaboration des valeurs organisationnelles peut devenir un enjeu de luttes politiques. Dans les années soixante-dix, d'après Robbins (1986: 473-74), nous sommes passés des normes et valeurs comme contraintes externes données face auxquelles l'organisation devait s'adapter, à une conception de l'organisation comme lieu de conflits

[114] Même le discours sur les profits se révèle creux pour Galbraith, car l'organisation complexe moderne ne poursuit pas les profits avant tout, mais bien la sécurité du revenu et la croissance (Galbraith 1968: 178).

politiques mettant en jeu ces normes. Un des principaux porte-parole de ce nouveau paradigme fut Jeffrey Pfeffer. Il établit une distinction nette entre le modèle *rationnel* où les buts organisationnels sont clairs et les actions entreprises au nom de l'atteinte de ces buts, et le modèle *politique* où les buts sont diffus et les acteurs tentent d'user de leur influence afin d'imposer leurs vues (Daft 1988: 417-18; Schermerhorn, Hunt, Osborn 1985: 560-63). Face à un enjeu capital dans une structure de pouvoir qui n'est pas complètement centralisée, si les acteurs influents expriment des intérêts divergents, alors seul le modèle politique s'avère pertinent (Pfeffer 1981: 70). Les agents ont leurs propres valeurs, et ils ont besoin de savoir que leurs actions ont une signification au sein de l'entreprise. L'activité politique consiste à fournir cette signification par la justification des relations de pouvoir et la rationalisation (au sens de "rendre rationnel") des décisions prises, sous la contrainte du respect des valeurs chères aux membres : "The task of those who benefit from these decisions is to legitimate and to justify them, to render power less visible and to provide justification for others acceding within the organization" (Pfeffer 1981: 182, aussi cf. 228-29).

Il est clair chez Pfeffer que les tentatives de légitimation et de justification se déroulent dans le champ du langage et du symbolisme. Deux facteurs contribuent à l'efficacité de ce type de représentation symbolique. D'abord, les agents s'habituent à la longue aux structures existantes de pouvoir et finissent par les considérer comme légitimes, voire même désirables (Pfeffer 1981: 5). Dans un tel cas, le pouvoir potentiel des employés, même ceux tout en bas de la pyramide, peut être élevé sans qu'ils ne s'en servent. L'ouvrier a le pouvoir d'entraver la production mais puisqu'il considère légitime l'autorité hiérarchique, cela ne lui vient pas à l'esprit[115]. Le second facteur est lié à l'incertitude (préférences ambivalentes, difficultés d'évaluation de la situation, information déficiente, etc.) et au manque d'intérêt pour les buts de l'organisation. L'agent qui ne comprend pas ce qui se passe, soit par ignorance ou parce que ça ne l'intéresse pas, peut se satisfaire d'une explication superficielle ou d'une redescription symbolique (Pfeffer 1981: 196-205). Le symbolisme peut également pallier à un manque de pouvoir, en faisant croire à l'agent que ses réalisations ont de l'importance. Pfeffer mentionne brièvement un cas très intéressant où un département s'est vu confier la tâche (symbolique) de

[115] Bell, Walker et Willer (2000: 140-41) critiquent le caractère trop "consensuel" de cette forme d'autorité légitime généralisée, et ils remarquent à juste titre que les formes plus directes de pouvoir (incitatifs pécuniaires, pouvoir de congédiement, etc.) sont ainsi laissés en plan.

rédiger des documents à l'intention d'un autre département afin de calmer le sentiment des employés de s'être fait transférer illégitimement leurs responsabilités. A la suite d'interviews, il remarque que ceux-ci ne savent pas vraiment si ces documents ont un effet sur les décisions du second département et que, surtout, *ils ne veulent pas le savoir* (Pfeffer 1981: 205). La délégation de tâche est tout ce qu'ils désirent. Malheureusement, son modèle du pouvoir légitime ne nous fournit pas de piste pouvant expliquer cette instance de *self-deception*.

La théorie du pouvoir de Crozier et Friedberg constitue un exemple de rationalité fonctionnelle telle que définie plus haut, soit une modélisation des relations de pouvoir strictement en termes d'intérêts et de stratégies, sans considérer la manière dont les agents instancient ces relations. Partant de la notion de "système d'action concret" comme champ structuré de relations humaines orienté autour d'une problématique, l'organisation est un type de système d'action concret qui se distingue par une formalisation de la structure dans un ensemble de règles concrètes (Crozier, Friedberg 1977: 286-87). Ce qui oriente les décisions dans une organisation, ce n'est pas la recherche d'efficacité, ni les "normes de comportement civilisé", mais bien la sauvegarde par l'agent de son influence (Crozier, Friedberg 1977: 324-25). Dans l'univers fortement réglé des organisations, le *locus* du pouvoir se retrouve dans les interstices des règles; le pouvoir de l'agent étant proportionnel à sa marge de liberté à l'intérieur de la structure[116] (Crozier, Friedberg 1977: 69). Le pouvoir est donc "le résultat toujours contingent de la mobilisation par les acteurs des sources d'incertitudes pertinentes qu'ils contrôlent dans une structure de jeu donné, pour leurs relations et tractations avec les autres participants à ce jeu" (Crozier, Friedberg 1977: 30). C'est une conceptualisation qui, selon Friedberg (1993: 117), dépasse les typologies habituelles en leur fournissant un mécanisme commun. Dans ce modèle, le but ultime des luttes de pouvoir semble n'impliquer que la marge de liberté, qui serait en quelque sorte sa propre fin. Ces luttes sont toutefois circonscrites par la nécessité de survie de l'organisation. L'exercice du pouvoir est perçu comme un échange. Le dominant se doit de satisfaire minimalement les attentes du dominé pour pouvoir "continuer le jeu"; en d'autres mots, pour que la manipulation réussisse, il doit se laisser manipuler à son

[116] "Et la force, la richesse, le prestige, l'autorité, bref, les ressources que possèdent les uns et les autres n'interviennent que dans la mesure où ils leur fournissent une liberté d'action plus grande" (Crozier, Friedberg 1977: 70).

tour (Crozier, Friedberg 1977: 104). En autant que la structure demeure relativement stable, le dominé accepte la relation s'il peut en tirer un avantage en retour.

Toute action individuelle entreprise dans le cadre du jeu est considérée comme une "stratégie rationnelle", quel qu'en soit le contenu intentionnel. La rationalité se voit déterminée par les contraintes du jeu, dans le sens où l'on peut toujours découvrir la rationalité de l'action en évaluant la structure du champ (Crozier, Friedberg 1977: 55-57). L'utilisation que les auteurs font de la notion de "stratégie" exclut la nécessité de recourir à la légitimité dans l'exercice du pouvoir. Dans l'étude du Monopole industriel (Crozier 1963), les ouvriers d'entretien jouissent d'un pouvoir sur les ouvriers de production, pourtant leurs égaux au plan hiérarchique, car ils contrôlent une source d'incertitude, les pannes de machines. Les ouvriers de production réagissent mal à cet état de fait, souvent de façon émotionnelle. Les ouvriers d'entretien se voient en quelque sorte forcés de modérer la pratique de leur pouvoir. Dans un modèle d'autorité légitime, l'indignation des ouvriers de production s'expliquerait par le caractère illégitime de cette relation de pouvoir violant la hiérarchie officielle. Crozier y voit plutôt une stratégie de leur part visant à restreindre le pouvoir des ouvriers d'entretien sur eux et donc à préserver leur propre liberté d'action. En autant que l'on puisse considérer une réaction émotionnelle comme une stratégie rationnelle, ce qui n'est pas du tout évident (voir Elster 1999), on s'aperçoit qu'il n'est jamais question de légitimité, mais seulement d'intérêts de chacune des parties. Le problème avec cette conception exclusivement instrumentale du pouvoir est qu'elle ne peut pas rendre compte des pratiques concrètes des agents qui elles, se voient la plupart du temps euphémisées.

Pour terminer, il existe en théorie des organisations un modèle inspiré de la théorie de l'échange social, la "théorie de la légitimation". La légitimité du pouvoir y est établie par deux critères, la validité (*validity*) qui est l'intelligibilité de la relation de pouvoir comme faisant partie d'un ordre des choses exogène et la convenance (*propriety*) qui est la croyance individuelle en la justesse de la relation (Bell, Walker, Willer 2000: 163). La validité ne met pas en jeu les valeurs des membres; une relation de pouvoir est considérée comme valide en autant qu'elle se fonde sur un ensemble de règles spécifiques (Thomas, Walker, Zelditch 1986: 380). Plusieurs sociologues des organisations perçoivent la convenance comme un effet de la

validité : les agents finissent par trouver juste une relation répétitive conforme aux règles. D'autres ont un point de vue divergent, à l'effet que la convenance ne peut provenir entièrement de la validité. Pour un subordonné, la validité rend seulement la relation "moins injuste" (Molm, Quist, Wiseley 1994: 103-4). La validité correspond assez bien à notre légitimité par référence à la structure formelle, et la convenance, à la légitimité dans sa variante informelle.

La structure formelle comme fondement de la légitimité

Au sein de la firme, on s'attend à ce que les dirigeants fassent preuve de rationalité dans leurs décisions en vue de l'atteinte des buts collectifs. Leur rôle est de coordonner les efforts des départements et des employés, et d'exiger que chacun d'eux atteigne ses objectifs. Un tel exercice de pouvoir se bute naturellement à la résistance des subordonnés lorsqu'on leur demande de fournir un effort supplémentaire, surtout si cette demande se situe au niveau personnel. La relation aura plus de chance d'être acceptée si elle correspond aux règles et aux procédures organisationnelles en vigueur auxquelles les subordonnés ont déjà accepté de se soumettre. Ainsi, pour que l'ordre puisse être entendu, le dirigeant devra dans bien des cas sacrifier la rationalité collective au nom de la routine : "Ce qui compte est de décider selon certaines pratiques considérées comme normales (habituelles et obligatoires pour un milieu donné), plus que la manière considérée comme la plus rationnelle dans une perspective économique, celle qui est censée guider l'action du management" (Alter 2003: 54-55). La structure formelle de l'organisation jouissant déjà d'une grande légitimité, elle devient une cible de choix pour la légitimation de l'autorité. Cette stratégie de légitimation aura toutefois des conséquences sur les pratiques sociales au sein de l'entreprise.

La structure formelle tend à s'autonomiser avec le temps. Initialement conçue au service des buts de l'entreprise, elle tend à n'être instrumentale qu'à son propre maintien. Les postes et les descriptions de tâches deviennent des chasse-gardées, la structure se sclérose par rapport à un environnement en mutation permanente, et la légitimité entre en conflit avec l'efficacité. Selon Meyer et Rowan (1992: 25), l'organisation typique est peuplée de "mythes de la structure formelle", où les règles en place s'avèrent plus importantes que l'efficacité comme critère d'allocation des tâches aux différents postes. La hiérarchie formelle, officielle, définit une structure impersonnelle des relations de pouvoir, reliée aux postes et non directement aux

individus occupant ces postes. L'exercice de pouvoir de poste à poste est plus facilement recevable qu'un exercice personnel car au moins les membres savent à quoi s'attendre. Les obligations et les frontières y sont habituellement clairement établies et connues de tous. Il s'agit d'un pouvoir *reconnu* car l'agent, à son entrée dans la firme, accepte de respecter la hiérarchie. Il fournira un effort supplémentaire lorsqu'on le lui demandera, en autant que les règles en la matière soient respectées, et il comprendra que ses supérieurs ont le pouvoir de le suspendre ou de le congédier s'il accomplit un travail médiocre. Ce pouvoir n'est toutefois pas *méconnu*, car il demeure perçu comme une authentique relation de pouvoir[117]. Simon (1983: 106) distingue entre agents loyaux aux buts de l'organisation et agents loyaux à l'organisation elle-même; si la poursuite des buts exige une restructuration, ce sera avec l'approbation des premiers et l'opposition des derniers. Nous pourrions toutefois ramener la loyauté à l'organisation à une loyauté plus locale, au sous-groupe, à la profession, ou au poste. Alter (2003) décrit une telle situation : bien que toutes sortes d'études démontrent que ce sont les ouvriers sur le plancher de l'usine qui savent le mieux comment optimiser le procédé de production, les dirigeants ne leur portent à peu près jamais attention, préférant déterminer "d'en haut" comment opérer les machines. Pour Alter, ceux-ci se privent de solutions rationnelles permettant d'augmenter la productivité au nom des normes sociales définissant les relations entre patrons et employés. Avant tout, "un décideur doit décider" (Alter 2003: 60). Cette norme s'oppose à l'idée qu'un décideur *puisse* décider, mais puisse également adopter les solutions des autres. Il faut en tout temps respecter la hiérarchie : "L'idée du sérieux est celle de la verticalité descendante des procédures" (Alter 2003: 61). Un dirigeant qui oserait demander aux ouvriers comment il faut faire, pour ensuite formaliser ces procédures, perdrait la face devant ses collègues de la direction; ce serait quelqu'un qui refuse de remplir correctement son rôle et qui transgresse la hiérarchie en conférant un droit de cité à ceux qui ne le "méritent" pas[118]. Ces deux maximes font bien plus qu'exprimer la primauté de la hiérarchie, elles constituent des

[117] Pour Meyer et Rowan, toute procédure qui ne se montre pas immédiatement efficace constitue un "rituel" ou un "mythe". C'est selon nous une exagération. Un comportement régulier a beau ne pas s'avérer efficace à chaque instant, il permet néanmoins de coordonner les attentes. Un investissement à long terme peut s'effectuer au détriment du court terme. Pourtant, des cas similaires sont analysés comme des instances d'inefficacité rituelle (Meyer, Rowan 1992: 37).

[118] Galbraith (1968) fait souvent référence à une norme hiérarchique semblable. Il est conscient que, bien que le pouvoir effectif soit passé du sommet vers la "technostructure" des experts, la croyance populaire en la pérennité de la hiérarchie pyramidale et au pouvoir suprême du PDG demeure, et ceci n'est pas sans conséquences sur les relations de travail.

normes sociales sur la façon acceptable de diriger, qui n'ont plus rapport avec l'efficacité organisationnelle. La validité formelle est ainsi intériorisée

Le poste confère un prestige social, particulièrement dans la société à l'extérieur de l'organisation. Comme en témoigne Galbraith:

> "(...) la grande société qui a atteint son plein stade de développement continue d'être le symbole du succès et de la réussite dans l'ordre de la civilisation. Elle communique ce prestige à ses membres : il vaut manifestement mieux appartenir à la General Motors ou à la Western Electric que d'être un isolé. La question que se posent automatiquement deux hommes qui font connaissance en Floride ou dans un avion est : 'Chez qui travaillez-vous ?' Tant que cette précision n'est pas connue, l'autre est une énigme; on ne peut pas le situer dans un contexte; personne ne sait quel degré de considération il mérite, pour ne pas parler de respect, ou même s'il est digne de la moindre attention" (Galbraith 1968: 162).

Nous pourrions ajouter, "Quel poste y occupez-vous ?". Dans un tel contexte, on peut exercer un pouvoir légitime en autant que l'on respecte les honneurs du poste. L'identification au poste (ou les attentes de rôle) et l'idée de professionnalisme représentent autant de besoins des membres allant souvent à l'encontre de la chaîne formelle de commandement. Homans révèle dans une enquête que les employées d'un bureau comptable tiennent absolument à maintenir leurs dossiers à jour même lorsque la direction ne l'exige pas. Il interprète ce phénomène comme un conflit entre la "logique de l'employé" et la "logique du management" (Homans 1962: 68), une réaction qu'il s'explique mal : "The members of a job group have an almost pathetic expectation that their boss should represent their interests and *help them behave according to their own norms* as against everybody else's norms, including management's" (Homans 1962: 69, italiques rajoutées).

La structure formelle prescrit également une *manière* de décider. Étant donné que les règles organisationnelles prescrivent les moyens appropriés en vue de la réalisation des buts de l'organisation, on s'attend que la prise de décision respecte en tout temps les canons de la rationalité. Chaque agent se doit de chercher à atteindre les buts de l'organisation de façon optimale, en faisant usage de données et de plans et en appliquant les méthodes de gestion en vigueur. Tout comme le poste, la pratique de la rationalité peut acquérir une indépendance par

rapport à son rôle dans la gestion efficace de l'entreprise et devenir une sorte de "religion" (Pfeffer 1981: 194). Nous avons vu plus haut la distinction que Pfeffer effectue entre le modèle rationnel et le modèle politique de décision. La légitimation qui s'opère ici est une euphémisation d'une décision de nature politique en décision rationnelle, donc une *rationalisation* de la décision. Le recours à la rationalité est une autre manière d'impersonnaliser la relation de pouvoir, en autant que cette rationalité soit *collective*, c'est-à-dire qu'elle se réfère aux buts de l'organisation. Pour Meyer (1992: 265), cette rationalité se pose comme un réseau (*map*) unifié des relations entre moyens et fins menant à la réalisation des buts de l'organisation. La rationalisation en termes de rationalité *individuelle* n'est pas acceptable, car elle signale une volonté d'accaparement des ressources de l'entreprise à des fins personnelles. Lorsque la décision rationnelle du point de vue de l'individu va à l'encontre de la rationalité organisationnelle prescrite, c'est cette dernière qui prévaut, et l'agent devra s'y adapter (Alter 2003: 34-35). On relève un tel conflit dans la promotion simultanée du "management participatif" qui cantonne les employés dans des objectifs établis et de l'"*empowerment*" qui les encourage à changer ces même objectifs en faisant preuve d'initiative personnelle (Bagla-Gökalp 1998: 98-99). Si l'on considère, d'un côté, que l'initiative personnelle comporte des risques pour l'agent et que la reconnaissance de ses réalisations par la direction demeure incertaine, et, de l'autre, que la soumission aux méthodes de gestion existantes représente une stratégie individuelle où le risque est quasi-nul et la reconnaissance prescrite par des règles et des normes, alors il y a de fortes chances que chacun se cantonne dans son rôle afin d'éviter les problèmes (Alter 2003: 241-43).

5.3 Conclusion

Les mécanismes de légitimation du pouvoir au sein d'une entreprise, selon les théories des organisations que nous avons abordées, portent surtout sur la structure formelle. Le conformisme à la structure organisationnelle est un pouvoir reconnu car il s'exprime dans des règles impersonnelles s'appliquant à tous et connues de tous, mais la relation de pouvoir n'est pas pour autant méconnue, pas, comme le prétendent Bachrach et Baratz, à la suite d'une argumentation raisonnable, mais bien parce que le contrat de travail semble établir une "zone de consentement" à la Simon, car la motivation à l'obéissance dépend en bout de ligne de la possibilité de sanctions positives ou négatives. L'autorité légitime, telle qu'entendue par les

théories des organisations, constitue une forme de légitimation externe partiellement méconnue. Les sanctions conservent leur force de motivation, mais l'agent dispose de raisons supplémentaires de se soumettre, comme le prestige du poste, la poursuite sincère des buts de l'organisation, la recherche du travail bien fait, etc. Ces raisons font en sorte que l'agent accepte plus facilement et plus "naturellement" la relation de pouvoir, sans pour autant croire qu'il agit entièrement en fonction de ses propres valeurs. L'euphémisation symbolique de Bourdieu se veut plus radicale. Ici l'agent méconnaît la relation de pouvoir; la possibilité de sanctions n'influence aucunement ses décisions. Il croit agir selon ses valeurs, et pour le bien de la communauté.

Le caractère légitime de la structure formelle dépend crucialement de son acceptation par tous les membres. En ce qui a trait au problème du pouvoir dans les organisations, la caractéristique première d'une entreprise par rapport à d'autres formes d'organisation est que l'agent y fait son entrée par un contrat explicite. L'admission au sein de l'entreprise représente un échange de force de travail pour une compensation pécuniaire. Si nous supposons en plus la liberté de choix du lieu où l'agent veut faire carrière, nous voyons bien qu'en partant, la relation entre l'agent et l'organisation équivaut à un échange marchand : l'agent accepte de se soumettre aux règles et à la hiérarchie en contrepartie d'un salaire. La première forme de soumission, fondamentale, est donc rationnelle et ouverte, contrairement aux relations masquées et (apparemment) non rationnelles du pouvoir symbolique. Le besoin de légitimation en organisation surgit de deux sources. D'abord, elle augmente l'efficacité de l'entreprise, car plus l'agent accepte l'autorité, plus il sera porté à s'auto-discipliner et l'organisation aura à dépenser moins de ressources en sanctions. Cela ne signifie pas pour autant que la promotion de la légitimité aura nécessairement pour conséquence une efficacité accrue; nous avons vu les effets pervers de l'autonomisation de la structure formelle et du fétichisme des règles qui surviennent lorsque les comportements doivent être redécrits en termes "officiels" pour être acceptés. L'autre facette du besoin de légitimité réside dans la réalisation de soi dans le travail. Bien souvent, l'employé ne se contente pas de son salaire; il a besoin en plus d'une satisfaction au travail et de la reconnaissance de ses pairs et de ses supérieurs. Il peut finir par s'attacher à son groupe de travail et ainsi à adhérer aux normes de ce groupe par opposition au reste de l'organisation.

Alors que dans la légitimation informelle, le besoin d'estime de soi et d'estime d'autrui constitue la motivation prédominante de l'agent à se soumettre, ces facteurs exercent une influence moindre au sein de l'entreprise. Le caractère fondamentalement rationnel, et reconnu comme tel, de la relation entre employé et employeur vient fausser la logique de Bourdieu de l'appartenance au groupe par la dénégation du comportement rationnel. Il faut dire que la règle organisationnelle se situe plus près d'une convention que d'une norme sociale[119]. Avec l'autorité légitime informelle, le dominant offre une représentation symbolique des valeurs des dominés, qu'ils ne connaissent qu'à un niveau pratique, alors qu'avec l'autorité légitime formelle, on peut mobiliser les dominés en invoquant des règles organisationnelles au caractère déjà objectif, qui ne sont pas sujettes à une telle représentation. Les règles sont ici exogènes; la légitimité de l'autorité peut s'avérer efficace même si l'agent n'exprime aucun attachement émotionnel à leur égard. On accepte ici un certain degré d'individualisme : bien que l'égoïsme pur ne soit pas convenable (au sens relatif qu'un comportement désintéressé sera toujours plus apprécié), personne n'exige que chacun se donne entièrement à l'entreprise. Personne n'est dupe du fait que chacun soit minimalement motivé à gagner de l'argent. Bien que le don de soi aux buts de l'organisation s'avère "payant" tout comme dans les stratégies de dévotion au groupe étudiées par Bourdieu, l'"intérêt au désintéressement" n'a pas besoin ici d'être tabou; chacun sait que l'on se sacrifie en bonne partie pour l'avancement. L'individualisme rationnel constitue dans ce champ particulier "un élément de la vie en commun" (Alter 2003: 226-7).

Tant qu'une structure formelle et impersonnelle est présente, la légitimation du pouvoir aura tendance à s'effectuer dans ce langage. Lorsque la structure formelle s'avère moins rigide, la légitimité se déplace vers les normes sociales et les valeurs du groupe. Les deux cas d'étude de Crozier (1963) se révèlent être des organisations à la réglementation très stricte, couvrant tous les aspects du procédé de travail, lui-même passablement routinier; il y relève très peu d'identification aux buts de l'organisation ou de luttes politiques, sauf dans des cas comme celui des ouvriers d'entretien. En particulier, les procédures automatiques d'avancement,

[119] Au sens où les règles sont perçues avant tout comme des moyens en vue de la réalisation d'un bien collectif, et non comme l'actualisation de valeurs à caractère social, comme le respect ou l'égalité. La différence principale est que dans la convention, l'action ouvertement rationnelle est considérée comme acceptable, alors que dans la norme sociale, une telle action sera perçue comme suspecte.

présentes dans les deux cas, ont sur les membres le même effet, soit la résignation et le manque d'intérêt dans les affaires de l'entreprise. Morrill (1991) a enquêté sur le comportement des cadres d'un fabricant de jouets avant et après une réforme managériale qui a eu pour effet d'assouplir les règles formelles et de conférer une plus grande liberté d'action aux cadres. Il a remarqué que, suite à la réforme, les relations se présentèrent beaucoup plus sous l'emprise de la logique du symbolisme et du code d'honneur. Lorsque les règles et les postes deviennent flous, il s'avère difficile d'euphémiser une relation de pouvoir par une impersonnalisation institutionnelle, alors celle-ci se découvre. Et lorsque en plus les distinctions entre les positions hiérarchiques s'estompent, la réputation des agents tend à remplacer la nomination formelle comme signe d'autorité, et les luttes de pouvoir prennent une tournure ouvertement politique. Lorsque la soumission rapporte à l'agent un capital matériel accepté *a priori* par le groupe, le rôle du capital symbolique dans l'acceptation de la soumission s'en trouve diminué. Galbraith reprend quelque peu cette idée lorsqu'il prétend que la raison qui pousse les cadres à être avant tout motivés par l'identification et l'adaptation est qu'ils gagnent suffisamment d'argent et qu'ils ne sont donc plus aussi motivés par une augmentation de salaire. Moins il est nécessaire de gagner sa vie au travail, donc, moins le besoin d'argent représente une contrainte structurelle, plus le membre demandera à être "payé" en capital symbolique de reconnaissance, et inversement, plus la contrainte financière pèse sur l'agent, moins il s'identifiera aux visées de ses supérieurs (Galbraith 1968: 141-43).

Le modèle du pouvoir symbolique s'avère plus pertinent dans des organisations où la reconnaissance constitue la motivation primordiale, et où les sanctions directes n'ont que peu d'effets, comme par exemple les associations politiques ou les organismes volontaires de bienfaisance. Bourdieu a lui-même appliqué son modèle à des organisations où le salaire se révèle soit sans importance, comme les partis politiques, soit méconnu, comme la bureaucratie française ou l'Église. Il a su toutefois relever la tension entre sanctions et symbolisme, comme dans cet exemple portant sur les partis politiques :

> "Plus le processus d'institutionnalisation du capital politique est avancé, plus la conquête des 'esprits' tend à se subordonner à la conquête des postes et plus les militants, liés par le seul dévouement à la 'cause', reculent au profit des 'prébendiers', comme les appelle Weber, sortes de clients, durablement liés à l'appareil par les

bénéfices et les profits qu'il leur assure, tenant à l'appareil pour autant que celui-ci les tient en leur redistribuant une part du butin matériel ou symbolique qu'il conquiert grâce à eux (...)" (Bourdieu 2001: 249-50).

CONCLUSION

A l'aide du modèle motivationnel que nous avons développé dans cette thèse, nous avons pu effectuer une analyse de certains champs sociaux avec des résultats originaux. Dans le champ de la délibération politique, nous avons vu que notre modèle permet ce que nous avons nommé la "délibération circonstancielle", soit une délibération à partir d'agents dont les motivations peuvent s'avérer instrumentales ou correspondant à des vertus civiques reliées à la décision ou à l'action de manière émotive. Selon les circonstances, ce type de délibération peut se montrer similaire aux modèles proposés par les théories de la démocratie délibérative. Comme approche du comportement politique, la délibération circonstancielle se situe entre deux pôles : la démocratie délibérative, qui considère le citoyen non rationnel ou "raisonnable", et les théories du choix social, qui se fonde sur le choix rationnel. Ensuite, dans le champ des normes sociales, nous avons distingué deux types de motivation au conformisme à la norme : le conformisme non rationnel, motivé par un respect des valeurs instanciées par la norme, et le conformisme rationnel, motivé par la maximisation d'utilité. Nous avons porté notre attention sur un cas particulier de conformisme rationnel, le conformisme hypocrite, lorsqu'il devient dans l'intérêt de l'agent de paraître motivé par les valeurs. L'équilibre hypocrite représente ce type d'équilibre qui peut survenir dans un groupe composé d'agents hypocrites où tous se conforment à la norme sans vraiment y croire. Nous avons prétendu qu'un tel équilibre est viable, quoique plus fragile qu'un équilibre "vertueux", et qu'il permet d'expliquer, entre autres, la persistance d'une norme néfaste pour la collectivité. L'étude de la pratique du duel nous a permis d'illustrer un tel phénomène social. A la lumière de notre modèle motivationnel, nous arrivons à mieux comprendre des pratiques qui semblent tout à fait étranges du seul point de vue de la rationalité. Finalement, nous avons appliqué notre modèle au champ du pouvoir au sein des organisations. Nous avons proposé un concept d'autorité légitime qui distingue entre la légitimation formelle, plus rationnelle et collée aux règles et à la hiérarchie officielle, et la légitimation informelle, issue des valeurs communes du groupe, mettant en cause par ce fait motivations rationnelles et non rationnelles. Sans avoir eu l'ambition de fournir une explication complète, nous avons tout de même pu mieux comprendre les différences de comportement en ce qui a trait à l'exercice de l'autorité dans des entreprises à la réglementation stricte par

rapport aux entreprises plus flexibles. Ainsi, on a pu observer un aspect de l'influence des institutions formelles sur les motivations des agents.

Nous avons proposé un modèle où les agents ne sont pas seulement disposés, mais *motivés* à agir de façon non rationnelle. En d'autres termes, certains *veulent* se distancer de la rationalité, à cause de certains principes éthiques, alors que d'autres *veulent paraître* s'en distancier, car cela est dans leur intérêt. Nous avons relevé plusieurs avantages d'un tel modèle par rapport à ceux de la théorie du choix rationnel. D'abord, l'explication des comportements, particulièrement lorsque des valeurs sont en jeu, s'avère plus réaliste. Dans de telles circonstances, la théorie du choix rationnel doit s'en tenir à des modélisations *ad hoc* des agents, qui les considèrent tous rationnels, afin de préserver la validité universelle de l'hypothèse de rationalité. Ces modèles doivent alors recourir à des hypothèses auxiliaires du comportement ou de la dynamique de la situation qui, bien souvent, ne correspondent pas aux intentions mêmes des agents. Comme nous l'avons souvent répété, cette approche n'est certainement pas dénuée d'intérêt en sciences sociales, mais nous croyons que l'on doive préférer une modélisation de l'agent qui correspond plus justement à la manière dont celui-ci se perçoit lui-même. La *folk psychology* de la théorie du choix rationnel, soit le complexe désirs-croyances-intentions, dont la simplicité et le pouvoir explicatif constituent ses principaux attraits en sciences sociales, peut être complétée par une *folk psychology* des valeurs, fondée sur les émotions de honte et de culpabilité. Un second avantage de notre modèle est la possibilité de rendre compte non seulement des comportements particuliers des agents qui ne sont pas motivés par la rationalité, mais aussi des effets de ces comportements sur l'action collective, particulièrement sur les agents rationnels. Nous avons vu que les "stratégies" des agents vertueux, qui n'en sont pas réellement car ceux-ci agissent par conviction plutôt que par intérêt calculé, façonnent l'éventail stratégique des agents rationnels, en déterminant comment il faut se comporter en société si l'on veut être respecté. Ce jeu des comportements convenables, sincères et hypocrites, ouvre la voie à une explication possible de nombre de comportements que l'on observe parfois dans les interactions sociales et qui ne nous apparaissent pas rationnels du premier coup d'oeil. Finalement, dans une perspective plus large, nous croyons que notre modèle offre la possibilité de rapprocher la théorie du choix rationnel, du moins dans sa version étendue, de théories sociales souvent en porte-à-faux avec

sa conception de l'action individuelle. Nous en avons tenté l'expérience avec la démocratie délibérative et la sociologie de Bourdieu; d'autres rapprochements fructueux demeurent tout à fait possibles.

Par contre, une approche non rationnelle de l'action collective, même liée de près à la théorie du choix rationnel, ne peut prétendre à la simplicité et à l'élégance des modèles rationnels. Un des principaux attraits de la théorie du choix rationnel, particulièrement dans sa version restreinte, ainsi qu'en théorie des jeux, est sa capacité de rendre compte de phénomènes sociaux complexes à l'aide de modèles quasi-mathématiques, et ainsi d'être en mesure de produire des explications aussi claires que possible. Un modèle prétendant que les agents ne calculent pas toujours, ou du moins que leurs calculs peuvent être biaisés par les sentiments, n'a d'autre choix que de devoir faire son deuil du caractère hautement analytique du cadre fourni par la théorie du choix rationnel, pour adopter à la place des mécanismes sociaux plus intuitifs. Certaines branches de la théorie du choix rationnel, comme le "*satisficing*" de Simon, la rationalité cognitive, et l'École de Chicago en économie, ont entrepris d'intégrer certains mécanismes de la psychologie béhavioriste, avec un certain succès. Comme nous nous intéressons plus particulièrement aux motivations des agents, nous avons choisi de nous référer à une psychologie plus classique des émotions, reliant les valeurs à l'estime de soi et à l'estime d'autrui.

En proposant cette thèse, notre but est de contribuer à une théorie de l'action collective en bonne et due forme, intégrant les motivations rationnelles et non rationnelles, qui reste à faire. Nous ne prétendons aucunement offrir une théorie complète de ce genre. Plusieurs éléments constitutifs d'une telle théorie restent à construire. Au niveau de la théorie des émotions, nous avons retenu des concepts très simples de honte et de culpabilité, dans le but explicite de demeurer dans le cadre de la *folk psychology* de la théorie du choix rationnel, qui se sert également de versions simplifiées de concepts mentaux comme les désirs, les croyances et l'intention. Nous avons proposé comme postulat que la honte et la culpabilité représentaient deux formes d'une même émotion globale, représentées sur un continuum d'intensité. Nous avons également supposé que les jugements portant sur soi et sur ses actions revenaient au même. Ce n'est pas là la seule manière d'aborder ces deux émotions. Un modèle motivationnel

similaire au nôtre pourrait se fonder sur une conception différente. D'autres émotions à caractère social pourraient également entrer en jeu, comme l'envie, la jalousie, la colère, la fierté, etc. Enfin, la relation que nous établissons entre émotions et valeurs est sujette à débats; de fait, on retrouve plusieurs propositions alternatives dans la littérature récente (Gibbard 1990; Tappolet 2002; Livet 2002). Les théories de l'idéologie en sociologie et en sciences politiques ont également le potentiel d'apporter une contribution significative à notre modèle. Il serait très instructif d'explorer les divers mécanismes de propagation des idées et des valeurs en société, et de l'influence de la réputation et des titres officiels sur la genèse et l'expression reconnue de ces idées et valeurs.

Au cours de nos recherches, nous avons exploré quelques domaines d'application de notre modèle. L'introduction de motivations non rationnelles dans une théorie rationnelle de l'action collective pourrait nous permettre de revisiter à profit plusieurs autres domaines de la vie sociale. En général, les instances d'action collective propices à une analyse de la sorte sont celles qui respectent les trois points suivants : les actions des individus ont un effet, réel ou prétendu, sur le bien-être collectif; ces actions sont publiques, au sens où elles sont observables[120], discernables et sanctionnables; et un certain niveau d'estime ou de réputation peut être gagné ou perdu selon les choix de chaque individu. Nous avons préalablement discuté de la démocratie délibérative; en fait, notre modèle s'appliquerait fort bien aux comportements à l'intérieur du champ politique en général, comme les discours, la gestion de l'image, les débats au Parlement, ou les campagnes électorales. Nous pourrions inclure également toute forme de négociation publique où un groupe défend à la fois ses valeurs et ses intérêts, comme dans les luttes syndicales et les revendications des groupes d'intérêt. Le modèle pourrait même nous éclairer sur la distinction entre négociations publiques et à huis clos. Enfin, nous pourrions rajouter tous les projets collectifs impliquant des valeurs communes, comme la sauvegarde de l'environnement, l'éducation, la protection des enfants, etc. En résumé, le modèle motivationnel que nous offrons est à son meilleur dans les situations d'action dites "politiques", au sens de l'expression "faire de la politique", qui peut tout aussi bien s'appliquer dans n'importe quel champ social. Faire de la politique, c'est se soucier de son image avant tout,

[120] Bien entendu, les *motivations* ne sont pas directement observables, sinon il n'y aurait pas de possibilité de comportement hypocrite.

surtout lorsque l'image d'un individu veillant à son propre intérêt ne vaut pas autant que celle de l'individu vertueux. Le paradoxe, c'est qu'on ne peut à la fois projeter une image désintéressée et une image de quelqu'un qui soigne son image. Pour étudier ce genre de comportement, il nous faut d'abord prendre au sérieux les motivations réellement désintéressées, qui sont la condition d'existence de la recherche de l'image, et ensuite élaborer ce que Bourdieu appelle "l'intérêt au désintéressement". Dans des champs politiques, familiaux, organisationnels, et autres où, bien souvent, les agents tirent satisfaction de leur réputation autant, sinon plus, que de l'atteinte de buts substantiels, un tel modèle pourrait jeter un éclairage nouveau sur bien des phénomènes sociaux.

BIBLIOGRAPHIE

- Alexander, Jeffrey C. (1995), *Fin de Siècle Social Theory*, Londres, Verso.
- Alter, Norbert (2003), *L'innovation ordinaire*, Paris, PUF, coll. "Quadrige".
- Anand, Paul (1987), "Are the Preferences Axioms Really Rational?", *Theory and Decision*, vol. 23, p. 189-214.
- Arnsperger, Christian (1998), "Engagement moral et optimisation individuelle", in Mahieu, F.-R.; Rapoport, H. (dir), *Altruisme. Analyses économiques*, Paris, Economica, p. 191-214.
- Audi, Robert (1985), "Self-Deception and Rationality", in Martin, M.W. (dir), *Self-Deception and Self-Understanding*, Univ. Press of Kansas, p. 169-93.
- Bachrach, Peter; Baratz, Morton S. (1970), *Power and Poverty*, New York, Oxford Univ. Press.
- Bagla-Gökalp, Lusin (1998), *Sociologie des organisations*, Paris, La découverte, coll. "Repères".
- Barrett, Karen C. (1995), "A Functionalist Approach to Shame and Guilt", in Tangney, J.P.; Fisher, K.W. (dirs), *Self-Conscious Emotions*, New York, Guilford Press, p. 25-63.
- Becker, Gary (1976), *The Economic Approach to Human Behavior*, Chicago Univ. Press.
- Becker, Gary (1986), "The Economic Approach to Human Behavior", in Elster, J. (dir), *Rational Choice*, New York Univ. Press, p. 108-122.
- Bell, Richard; Walker, Henry A.; Willer, David (2000), "Power, Influence, and Legitimacy in Organizations", in Bachrach, S. B.; Lawler, E. J. (dir), *Research in the Sociology of Organizations*, vol. 17, "Organizational Politics", p. 131-177.
- Benhabib, Seyla (1996), "Toward a Deliberative Model of Democratic Legitimacy", in Benhabib, S. (dir), *Democracy and Difference*, Princeton, Princeton Univ. Press.
- Ben-Ner, Avner; Putterman, Louis (1998), "Values and Institutions in Economic Analysis", in Ben-Ner, A.; Putterman, L. (dir), *Economics, Values, and Organization*, New York, Cambridge Univ. Press, p. 3-72.
- Bernheim, B. Douglas (1994), "A Theory of Conformity", *Journal of Political Economy*, vol. 102, p. 841-877.
- Bicchieri, Cristina (1993), *Rationality and Cooperation*, New York, Cambridge Univ. Press.
- Billacois, François (1986), *Le duel dans la société française des XVIe-XVIIe siècles*, Paris, EHESS.

- Binmore, Ken (1987), "Modeling Rational Players. Part I", *Economics and Philosophy*, vol. 3, p. 179-214.
- Boudon, Raymond (1998), "Social Mechanisms Without Black Boxes", in Hedström, P; Swedberg, R. (dir), *Social Mechanisms*, New York, Cambridge Univ. Press, p. 172-203.
- Boudon, Raymond (2003a), *Raison, bonnes raisons*, Paris, PUF.
- Boudon, Raymond (2003b), "Rationalité", in Boudon, R. *et al.* (dir), *Dictionnaire de sociologie*, Paris, Larousse, p. 194-195.
- Bourdieu, Pierre (1980), *Le sens pratique*, Paris, Minuit.
- Bourdieu, Pierre (1987), *Choses dites*, Paris, Minuit.
- Bourdieu, Pierre (1994), *Raisons pratiques*, Paris, Seuil, coll. "Essais".
- Bourdieu, Pierre (2001), *Langage et pouvoir symbolique*, Paris, Seuil, coll. "Essais".
- Bourdieu, Pierre (2003), *Méditations pascaliennes*, Paris, Seuil, coll. "Essais".
- Bowles, Samuel; Gintis, Herbert (1998), "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, p. 206-230.
- Brennan, Goeffrey; Pettit, Philip (2004), *The Economy of Esteem*, New York, Oxford Univ. Press.
- Brennan, Timothy J. (1989), "A Methodological Assessment of Multiple Utility Frameworks", *Economics and Philosophy*, vol. 5, p. 189-208.
- Brennan, Timothy J. (1991), "The Trouble with Norms", in Koford, K.J.; Miller, J.B. (dir), *Social Norms and Economic Institutions*, Ann Arbor, Univ. of Michigan Press, p. 85-94.
- Chazel, François (1983), "Pouvoir, structure et domination", *Revue française de sociologie*, vol. 24, p. 369-393.
- Christiano, Thomas (1993), "Social Choice and Democracy", in Copp, D.; Hamilton, J.; Roemer, J. (dir), *The Idea of Democracy*, New York, Cambridge Univ. Press.
- Cohen, Joshua (1986), "An Epistemic Conception of Democracy", *Ethics*, vol. 97, p. 26-38.
- Cohen, Joshua (1989), "Deliberation and Democratic Legitimacy", in Hamlin, A. (dir), *The Good Polity*, Oxford, Blackwell.
- Cohen, Joshua (1996), "Procedure and Substance in Deliberative Democracy", in Benhabib, S. (dir.), *Democracy and Difference*, Princeton, Princeton Univ. Press.
- Coleman, James S. (1986), *Individual Interests and Collective Action*, New York, Cambridge Univ. Press.

- Coleman, James S. (1987), "Norms as Social Capital", in Radnitzky, G.; Bernholz, P. (dir), *Economic Imperialism*, New York, Paragon House, p. 133-155.
- Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge (MA), Belknap Press.
- Cooke, Maeve (2000), "Five Arguments for Deliberative Democracy", *Political Studies*, vol. 48, p. 947-69.
- Cowen, Tyler (2002), "The Esteem Theory of Norms", *Public Choice*, vol. 113, p. 211-224.
- Cox, Andrew; Furlong, Paul; Page, Edward (1985), *Power in Capitalist Societies*, Brighton, Wheatsheaf.
- Crozier, Michel; Friedberg, Erhard (1977), *L'acteur et le système*, Paris, Seuil, coll. "Essais".
- Crozier, Michel (1963), *Le phénomène bureaucratique*, Paris, Seuil, coll. "Points".
- Daft, Richard L. (1988), *Organization Theory and Design*, 3e éd., St Paul (MN), West.
- Dahl, Robert A. (1991), *Modern Political Analysis*, 5e éd., Englewood Cliffs (NJ), Prentice Hall.
- Danielson, Peter (2004), "Rationality and Evolution", in Mele, A.R. (dir), *The Oxford Handbook of Rationality*, Oxford Univ. Press, p. 417-437.
- Demeulenaere, Pierre (1996), *Homo oeconomicus*, Paris, PUF.
- Dillon, Robin S. (1997), "Self-Respect: Moral, Emotional, Political", *Ethics*, vol. 107, p. 226-49.
- Elster, Jon (1983), *Sour Grapes*, New York, Combridge Univ. Press.
- Elster, Jon (1986a), "Introduction", in Elster, J. (dir), *Rational Choice*, New York Univ. Press, p. 1-33.
- Elster, Jon (1986b), "The Market and the Forum", in Elster, J.; Hylland, A. (dir), *Foundations of Social Choice Theory*, New York, Cambridge Univ, Press.
- Elster, Jon (1989a), *Solomonic Judgments*, New York, Cambridge Univ. Press.
- Elster, Jon (1989b), *The Cement of Society*, New York, Cambridge Univ. Press.
- Elster, Jon (1994a), "Argumenter et négocier dans deux Assemblées constituantes", *Revue française de science politique*, vol. 44, p. 187-256.
- Elster, Jon (1994b), "Rationality, Emotions, and Social Norms", *Synthese*, vol. 98, p. 21-49.

- Elster, Jon (1995), "Rationalité et normes sociales : un modèle pluridisciplinaire", in Gérard-Varet, L.; Passeron, J.-C. (dirs), *Le modèle et l'enquête*, Paris, Éd. de l'ÉHESS, p. 139-48.
- Elster, Jon (1998a), "A Plea for Mechanisms", in Hedström, P; Swedberg, R. (dir), *Social Mechanisms*, New York, Cambridge Univ. Press, p. 45-73.
- Elster, Jon (1998b), "Deliberation and Constitution Making", in Elster, J. (dir), *Deliberative Democracy*, Ney York, Cambridge Univ. Press, p. 97-122.
- Elster, Jon (1999), *Alchemies of the Mind*, New York, Cambridge Univ. Press.
- Elster, Jon (2000), *Ulysses Unbound*, New York, Cambridge Univ. Press.
- Elster, Jon (2004), "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, p. 3-14.
- Etzioni, Amitai (1986), "The Case for a Multiple-Utility Conception", *Economics and Philosophy*, vol. 2, p. 159-183.
- Etzioni, Amitai (1988), *The Moral Dimension*, New York, Free Press.
- Ferejohn, John (2002), "Rational Choice Theory and Social Explanation", *Economics and Philosophy*, vol. 18, p. 211-234.
- Follesdal, Dagfinn (1982), "The Status of Rationality Assumptions in Interpretation and in the Explanation of Action", *Dialectica*, vol. 36, p. 301-316.
- Frank, Robert H. (1998), "Social Norms as Positional Arms Control Agreements", in Ben-Ner, A.; Putterman, L. (dir), *Economics, Values, and Organization*, New York, Cambridge Univ. Press, p. 275-295.
- Friedberg, Erhard (1993), *Le pouvoir et la règle*, Paris, Seuil.
- Galbraith, John Kenneth (1968), *Le nouvel État industriel*, Paris, Gallimard.
- Galbraith, John Kenneth (1983), *The Anatomy of Power*, Boston, Houghton Mifflin.
- Gibbard, Alan (1990), *Wise Choices, Apt Feelings*, Cambridge (MA), Harvard Univ. Press.
- Goetschy, Janine (1981), "Les théories du pouvoir", *Sociologie du travail*, vol. 23, p. 444-467.
- Goldthorpe, John H. (2000), *On Sociology*, New York, Oxford Univ. Press.
- Goodin, Robert E. (1986), "Laundering Preferences", in Elster, J.; Hylland A. (dir), *Foundations of Social Choice Theory*, New York, Cambridge Univ. Press, p. 75-102.
- Goodin, Robert E. (1996), "Institutions and Their Design", in Goodin, R.E. (dir), *The Theory of Institutional Design*, New York, Cambridge Univ. Press, p. 1-53.

- Greenspan, Patricia (1993), "Guilt as an Identificatory Mechanism", *Pacific Philosophical Quarterly*, vol. 74, p. 46-59.
- Gutmann, Amy; Thompson, Dennis (1995), "Moral Disagreement in a Democracy", *Social Philosophy and Policy*, vol. 12, p. 87-110.
- Gutmann, Amy; Thompson, Dennis (1996), *Democracy and Disagreement*, Cambridge (MA), Belknap.
- Gutmann, Amy; Thompson, Dennis (2000), "Why Deliberative Democracy is Different", *Social Philosophy and Policy*, vol. 17, p. 161-180.
- Halliday, Hugh A. (1999), *Murder Among Gentlemen*, Toronto, Robin Brass Studio.
- Hardin, Russel (1995), *One for All: The Logic of Group Conflict*, Princeton Univ. Press.
- Harsanyi, John C. (1990), "Advances in Understanding Rational Behavior", in Moser, P.K. (dir), *Rationality and Action*, New York, Cambridge Univ. Press, p. 271-293.
- Hausman, Daniel M. (1992), *The Inexact and Separate Science of Economics*, Cambridge (UK), Cambridge Univ. Press.
- Hausman, Daniel M.; McPherson, Michael S. (1993), "Taking Ethics Seriously", *Journal of Economic Literature*, vol. 31, p. 671-731.
- Heath, Joseph (2001), "Rational Choice with Deontic Constraints", *Canadian Journal of Philosophy*, vol. 31, p. 361-388.
- Heckathorn, Douglas D. (1988), "Collective Sanctions and the Creation of Prisoner's Dilemma Norms", *American Journal of Sociology*, vol. 94, p. 535-562.
- Heckathorn, Douglas D. (1990), "Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control", *American Sociological Review*, vol. 55, p. 366-384.
- Hedström, Peter; Swedberg, Richard (dir), *Social Mechanisms*, New York, Cambridge Univ. Press.
- Hirschleifer, Jack (1987), "On the Emotions as Guarantors of Threats and Promises", in Dupré, J. (dir), *The Latest on the Best: Essays on Evolution and Optimality*, Cambridge (MA), MIT Press, p. 307-326.
- Höllander, Heinz (1990), "A Social Exchange Approach to Voluntary Cooperation", *The American Economic Review*, vol. 80, p. 1157-1167.
- Hollis, Martin (1996), *Reason in Action*, New York, Cambridge Univ. Press.
- Homans, George C. (1958), "Social Behavior as Exchange", *American Journal of Sociology*, vol. 63, p. 597-606.
- Homans, George C. (1962), *Sentiments and Activities*, New York, Glencoe.

- Isaac, Alan G. (1997), "Morality, Maximization, and Economic Behavior", *Southern Economic Journal*, vol. 63, p. 559-570.
- Jacoby, Mario (1994), *Shame and the Origins of Self-Esteem*, Londres, Routledge.
- Johnson, James (1998), "Arguing for Deliberation: Some Skeptical Considerations", in Elster, J. (dir), *Deliberative Democracy*, New York, Cambridge Univ. Press.
- Klaassen, Johann A. (2001), "The Taint of Shame: Failure, Self-Distress, and Moral Growth", *Journal of Social Philosophy*, vol. 32, p. 174-96.
- Knight, Jack; Ensminger, Jean (1998), "Conflict over Changing Social Norms: Bargaining, Ideology, and Enforcement", in Nee, V.; Brinton, M.C. (dir), *The New Institutionalism in Sociology*, New York, Russel Sage Foundation, p. 105-126.
- Knight, Jack; Johnson, James (1994), "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy", *Political Theory*, vol. 22, p. 277-296.
- Kolm, Serge-Christophe (1986), *Philosophie de l'économie*, Paris, Seuil.
- Kuran, Timur (1995), *Private Truths, Public Lies*, Cambridge (MA), Harvard Univ. Press.
- Kuran, Timur (1998), "Moral Overload and its Alleviation", in Ben-Ner, A.; Putterman, L. (dir), *Economics, Values, and Organization*, New York, Cambridge Univ. Press, p. 231-266.
- Lazar, Ariela (1999), "Deceiving Oneself Or Self-Deceived? On the Formation of Beliefs 'Under the Influence'", *Mind*, vol. 108, p. 265-290.
- Levitt, Steven D.; Dubner, Stephen J. (2005), *Freakonomics*, New York, William Morrow.
- Lewis, Helen Block (1971), *Shame and Guilt in Neurosis*, New York, International Universities Press.
- Lewis, Michael (1993), "Self-Conscious Emotions: Embarrassment, Pride, Shame and Guilt", in Lewis, M.; Haviland, J.M. (dirs), *Handbook of Emotions*, New York, Guilford Press, p. 563-73.
- Livet, Pierre (2001), "Action et cognition en sciences sociales", in Berthelet, J.-M. (dir), *Épistémologie des sciences sociales*, Paris, PUF, p. 269-316.
- Livet, Pierre (2002), *Émotions et rationalité morale*, Paris, PUF.
- Luban, David (1996), "The Publicity Principle", in Goodin, R. (dir), *The Theory of Institutional Design*, New York, Cambridge Univ. Press.
- Lukes, Steven (1977), *Essays in Social Theory*, Londres, Macmillan.
- Manion, Jennifer C. (2002), "The Moral Relevance of Shame", *American Philosophical Quarterly*, vol. 39, p. 73-90.

- Margolis, Howard (1990), "Dual Utilities and Rational Choice", in Mansbridge, J.J. (dir), *Beyond Self-Interest*, Univ. of Chicago Press, p. 239-253.
- Margolis, Howard (1991), "Incomplete Coercion: How Social Preferences Mix with Private Preferences", in Monroe, K.R. (dir), *The Economic Approach to Politics*, New York, Harper Collins, p. 353-370.
- Margolis, Howard (2003), "Cognition and Extended (NSNX) Rational Choice: Some Early Results", Working paper, Harris School of Public Policy, Univ. of Chicago.
- McAdams, Richard H. (1997), "The Origin, Development, and Regulations of Norms", *Michigan Law Review*, vol. 96, p. 338-433.
- Meyer, John W.; Rowan, Brian (1992), "Institutionalized Organizations: Formal Structure as Myth and Ceremony", in Meyer, J. W.; Scott, R. W. (dir), *Organizational Environments: Ritual and Rationality*, Newbury Park (CA), Sage, p. 21-44.
- Meyer, John W. (1992), "Conclusion: Institutionalization and the Rationality of Formal Organizational Structure", in Meyer, J. W.; Scott, R. W. (dir), *Organizational Environments: Ritual and Rationality*, Newbury Park (CA), Sage, p. 261-282.
- Miceli, Maria; Castelfranchi, Cristiano (1998), "How to Silence One's Conscience: Cognitive Defenses Against the Feeling of Guilt", *Journal for the Theory of Social Behaviour*, vol. 28, p. 287-318.
- Miller, David (1992), "Deliberative Democracy and Social Choice", *Political Studies*, vol. 40, p. 54-67.
- Molm, Linda D.; Quist, Theron M.; Wiseley, Phillip A. (1994), "Imbalanced Structures, Unfair Strategies: Power and Justice in Social Exchange", *American Sociological Review*, vol. 59, p. 98-121.
- Morrill, Calvin (1991), "Conflict Management, Honor, and Organizational Change", *American Journal of Sociology*, vol. 97, p. 585-621.
- Nagel, Jack H. (1975), *The Descriptive Analysis of Power*, New Haven, Yale Univ. Press.
- Nee, Victor; Ingram, Paul (1998), "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, p. 19-45.
- Ogien, Ruwen (2002), *La honte est-elle immorale ?*, Paris, Bayard.
- Ostrom, Elinor (1986), "An Agenda for the Study of Institutions", *Public Choice*, vol. 48, p. 3-25.
- Ostrom, Elinor (1998), "A Behavioral Approach to the Rational Choice Theory of Collective Action", *American Political Science Review*, vol. 92, p. 1-22.
- Pettit, Philip (1990), "*Virtus Normativa*: Rational Perspectives", *Ethics*, vol. 100, p. 725-755.

- Pettit, Philip (1995), "The Virtual Reality of *Homo Economicus*", *The Monist*, vol. 78, p. 308-329.
- Pettit, Philip (2001), "Deliberative Democracy and the Dircursive Dilemma", Australian National University, Social and Political Theory Working Paper W11.
- Pfeffer, Jeffrey (1981), *Power in Organizations*, Marshfield (MA), Pitman.
- Quéré, Louis (1995), "Le schématisme de la norme d'un point de vue sociologique", *Cahiers de philosophie politique et juridique*, no. 27, p. 227-251.
- Rabin, Matthew (1998), "Psychology and Economics", *Journal of Economic Literature*, vol. 36, p. 11-46.
- Raven, Bertram H. (1965), "Social Influence and Power", in Steiner, I.D.; Fishbein, M. (dir), *Current Studies in Social Psychology*, New York, Holt, Rinehart and Wilson, p. 371-382.
- Rawls, John (1971), *A Theory of Justice*, Cambridge (MA), Belknap Press.
- Rawls, John (1995), *Libéralisme politique*, Paris, PUF.
- Robbins, Stephen P. (1986), *Organization Theory*, Englewood Cliffs (NJ), Prentice Hall.
- Rosenberg, Alexander (1992), *Economics: Mathematical Politics or Science of Diminishing Returns?*, Univ. of Chicago Press.
- Ruffle, Bradley J. (1999), "Gift-Giving with Emotions", *Journal of Economic Behavior and Organization*, vol. 39, p. 399-420.
- Sabini, John; Silver, Maury (1997), "In Defense of Shame: Shame in the Context of Guilt and Embarrassment", *Journal of the Theory of Social Behaviour*, vol. 27, p. 1-15.
- Satz, Debra; Ferejohn, John (1994), "Rational Choice and Social Theory", *Journal of Philosophy*, vol. 91, p. 71-87.
- Schelling, Thomas C. (1960), *The Strategy of Conflict*, Cambridge (MA), Harvard Univ. Press.
- Schermerhorn, John R.; Hunt, James G.; Osborn, Richard N. (1985), *Managing Organizational Behavior*, New York, John Wiley & Sons.
- Schmidt, Christian (2001), *La théorie des jeux. Essai d'interprétation*, Paris, PUF.
- Schwartz, Warren F.; Baxter, Keith; Ryan, David (1984), "The Duel: Can These Gentlemen Be Acting Efficiently ?", *Journal of Legal Studies*, vol. 13, p. 321-355.
- Scott-Kakures, Dion (1996), "Self-Deception and Internal Irrationality", *Philosophy and Phenomenological Research*, vol. 56, p. 31-56.
- Sen, Amartya (1982), "Rational Fools", in Sen, A. (dir), *Choice, Welfare, and Measurement*, Cambridge (MA), MIT Press, p. 84-106.

- Sen, Amartya (1986), "Foundations of Social Choice Theory: An Epilogue", in Elster, J.; Hylland, A. (dir), *Foundations of Social Choice Theory*, New York, Cambridge Univ, Press.
- Shepsle, Kenneth A. (1989), "Studying Institutions: Some Lessons From the Rational Choice Approach", *Journal of Theoretical Politics*, vol. 1, no. 2, p. 131-47.
- Simon, Herbert A.; et al. (1992), Decision Making and Problem Solving", in Zey, M. (dir), *Decision Making*, Newbury Park (CA), Sage, p. 32-53.
- Simon, Herbert A. (1978), "Rationality as Process and as Product of Choice", *American Economic Review*, vol. 68, p. 1-16.
- Simon, Herbert A. (1983), *Administration et processus de décision*, Paris, Economica.
- Snyder, C.R. (1985), "Collaborative Companions: The Relationship of Self-Deception and Excuse Making", in Martin, M.W. (dir), *Self-Deception and Self-Understanding*, Univ. Press of Kansas, p. 35-51.
- Statman, Daniel (1997), "Hypocrisy and Self-Deception", *Philosophical Psychology*, vol. 10, p. 57-71.
- Statman, Daniel (2000), "Humiliation, Dignity and Self-Respect", *Philosophical Psychology*, vol. 13, p. 523-540.
- Stigler, George J.; Becker, Gary S. (1977), "De Gustibus Non Est Disputandum", *The American Economic Review*, vol. 67, p. 76-90.
- Sunstein, Cass R. (1993), "Democracy and Shifting Preferences", in Copp, D.; Hamilton, J.; Roemer, J. (dir), *The Idea of Democracy*, New York, Cambridge Univ. Press.
- Swartz, David (1997), *Culture and Power: The Sociology of Pierre Bourdieu*, Univ. of Chicago Press.
- Tan, Tommy C.; Werlang, Sergio R.D.C. (1988), "A Guide to Knowledge in Games", in Vardi, M.Y. (dir), *Proceedings on the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, Los Altos (CA), Morgan Kaufmann, p. 163-177.
- Tangney, June Price; et al. (1996), "Are Shame, Guilt, and Embarrassment Distinct Emotions?", *Journal of Personality and Social Psychology*, vol. 70, p. 1256-69.
- Tappolet, Christine (2002), *Émotions et valeurs*, Paris, PUF.
- Thomas, George M.; Walker, Henry A.; Zelditch, Morris jr. (1986), "Legitimacy and Collective Action", *Social Forces*, vol. 65, p. 378-404.
- van den Berg, Axel (1998), "Is Sociological Theory Too Grand for Social Mechanisms?", in Hedström, P.; Swedberg, R. (dir), *Social Mechanisms*, New York, Cambridge Univ. Press.

- Voirol, Olivier (2004), "Reconnaissance et méconnaissance: sur la théorie de la violence symbolique", *Information sur les sciences sociales*, vol. 43, p. 403-433.
- Walliser, Bernard (1989), "Instrumental Rationality and Cognitive Rationality", *Theory and Decision*, vol. 27, p. 7-36.
- Weibull, Jörgen W. (1998), "Evolution, Rationality and Equilibrium in Games", *European Economic Review*, vol. 42, p. 641-649.
- Wells, C.A. Harwell (2001), "The End of the Affair? Anti-Dueling Laws and Social Norms in Antebellum America", *Vanderbilt Law Review*, vol. 54, p. 1805-1847.
- Wolfelsperger, Alain (2001), "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, p. 63-92.
- Wrong, Dennis H. (1988), *Power, Its Forms, Bases, and Uses*, Univ. of Chicago Press.
- Young, Iris Marion (1996), "Communication and the Other: Beyond Deliberative Democracy", in Benhabib, S. (dir), *Democracy and Difference*, Princeton, Princeton Univ. Press.
- Zelizer, Viviana A. R. (1994), *The Social Meaning of Money*, New York, Basic Books.