

Les préférences sociales peuvent-elles constituer les prémisses d'une théorie positive de la morale en économie ?

Sylvie Thoron
Université Paris-Est, LIPHA, UPEC
Sylvie.thoron@u-pec.fr

Sylvie Thoron est Professeur d'économie à l'Université Paris-Est Créteil et membre du LIPHA. Après avoir longtemps consacré sa recherche aux interactions sociales, à la théorie des jeux, à l'économie publique et à l'économie de l'environnement, elle adopte aujourd'hui une approche plus épistémologique et s'intéresse en particulier aux modèles de comportement qui fondent les sciences économiques.

Article référencé comme suit :

Thoron, S. (2016) "Les préférences sociales peuvent-elles constituer les prémisses d'une théorie positive de la morale en économie ?" in *Ethique. La vie en question*, mai 2016.

Introduction

On aurait pu penser que la science économique dominante était parvenue à évacuer la morale de son champ d'analyse. Les questions de justice sociale et de justice distributive se trouvaient cantonnées dans la théorie du choix social, une approche normative qui, bien que respectée dans le monde académique, n'avait aucune influence sur le reste de la profession et les recommandations que celle-ci pouvait adresser aux politiques. L'exemple de l'influence d'Amartya Sen dans les instances internationales ne faisant, selon nous, que confirmer la règle. La science économique dominante se justifiait de cette mise à l'écart des questions liées à la morale en évoquant au moins trois raisons. Premièrement, parce que, prenant comme référence fondamentale le libéralisme économique de *La Richesse des nations*, elle pouvait stipuler avec Adam Smith que le marché n'avait pas besoin de morale. Deuxièmement, parce que l'impossibilité de la comparaison interpersonnelle des utilités constituait une justification technique. Enfin, parce qu'elle semblait avoir trouvé dans l'optimum de Pareto le critère qui lui permettait de comparer et juger les situations sociales et donc de rester une science de la décision publique. Ainsi, semblait-il y avoir un quasi-consensus à ce sujet. Non pas que la science économique ait jamais été parfaitement homogène mais les controverses entre

économistes ne portaient pas sur la moralité mais sur l'efficacité du marché. Sourds aux attaques des autres sciences sociales, de la philosophie, du politique et du profane¹, lorsque les uns et les autres parlaient de l'immoralité des marchés, les économistes semblaient attribuer ces critiques à une incompréhension des fondements du libéralisme économique et à une confusion entre immoralité et irrationalité ou inefficacité.

Ce quasi-consensus est cependant devenu de plus en plus fragile au tournant de notre XXI^{ème} siècle. L'économie expérimentale et comportementale avait depuis longtemps montré l'incapacité du modèle de l'agent économique à reproduire les comportements observés dans le laboratoire et dans la réalité économique. Dans un premier temps cependant, l'interprétation alternative fut souvent celle de la rationalité bornée des agents, jusqu'à ce que l'interprétation à travers le prisme de leur moralité s'impose finalement. Autre signe de ce retour de la morale en économie, Bruni et Sugden (2013) veulent aujourd'hui répondre aux accusations des philosophes et invitent les économistes à réfléchir aux vertus du marché.

Ainsi, il semblerait que l'économie ait aujourd'hui besoin d'une théorie positive de la morale. Par théorie positive de la morale nous entendons une théorie qui pourrait expliquer comment ce que l'on désigne sous le terme générique de « morale » émerge et façonne le comportement des individus en société. Il ne s'agit donc pas d'une théorie normative mais bien d'une théorie qui pourrait compléter les théories du comportement en économie.

Une théorie positive de la morale n'existe-t-elle donc pas déjà ? L'économie expérimentale et comportementale ne s'est pas limitée à faire la preuve d'observations contrariantes pour le modèle de l'agent économique. Elle a apporté des éléments d'interprétation et de modélisation. La théorie des préférences sociales a permis de modifier le modèle de la fonction d'utilité pour lui permettre de reproduire les comportements pro-sociaux observés dans les expériences comme des faits stylisés. La théorie des jeux évolutionnaires a fourni des modèles permettant d'expliquer l'apparition de comportements moraux ou pro-sociaux comme résultat d'un processus de sélection. Nous nous proposons ici de faire un bref état des lieux d'une théorie positive de la morale en économie. L'objectif ne sera pas d'être exhaustif mais plutôt de mettre en évidence les difficultés qu'une telle entreprise présente pour la

¹ Pour s'en convaincre on pourra lire, par exemple, les attaques que Alasdair MacIntyre adresse au marché et aux économistes dans *After virtue* (1981), Didier Fassin et Laurence Fontaine sociologues et anthropologues qui reprennent le concept d'économie morale de W. Thomson, le discours du Bourget de François Hollande pendant la campagne de 2012...

science économique et les autres sciences du comportement. La réintroduction de la morale en économie, si elle a lieu, ne se fera pas à moindre frais.

Nous montrerons que la volonté de préserver le cadre théorique a jusque-là souvent primé sur la nécessité d'expliquer les comportements réels. Or, l'individualisme méthodologique inhérent à ce cadre rend difficile la compréhension de ce que l'on peut désigner de façon générale sous le terme de *mécanismes*, c'est-à-dire de ce qui, dans l'environnement des sujets et leur interaction pourrait expliquer les comportements pro-sociaux ou moraux. Nous soutiendrons, a contrario, que seule une approche résolument axée sur ces mécanismes peut ouvrir la voie vers une théorie positive de la morale en économie et nous défendrons ici cette approche, quitte à ce que cela nécessite de sortir du cadre de la théorie de l'utilité. Considérant que cette démarche est celle déjà adoptée par Adam Smith dans *la Théorie des sentiments moraux*, nous tenterons de savoir si l'économie expérimentale, aidée des autres sciences du comportement et notamment des neurosciences sociales, peuvent y apporter un éclairage nouveau et complémentaire.

1. *Une théorie positive de la morale existe-t-elle déjà ?*

Dans cette première partie, nous nous intéressons à une littérature expérimentale qui pourrait sembler servir de prémisse à une théorie positive de la morale. Nous ne tenterons pas ici d'en donner une description exhaustive, nous voulons seulement mettre en évidence les tournants importants de son développement. En particulier, nous ferons l'impasse sur les expériences liées à la théorie du choix social qui tentent d'éliciter les principes de justice tels qu'ils sont perçus dans la population et qui reposent essentiellement sur des protocoles à vignette ou questionnaire. La méthode de ce type d'expérience est de décrire une situation aussi réelle que possible et de demander à des sujets de proposer un partage qu'ils considèrent comme juste. La plupart de ces expériences ne comportent pas d'incitations pour les sujets. Pour toute cette littérature on pourra se référer en particulier à une longue revue de la littérature de Konow (2003) et à un ouvrage sur ce thème de Gaertner et Schokkaert (2012). Nous délaierons ici cette littérature car elle a pour objectif *une analyse positive des théories de la justice* et de la morale pour reprendre le titre de l'article de Konow, alors que nous recherchons ici les prémisses d'*une théorie positive de la morale*. Dans la première approche les expériences servent à tester la validité empirique des principes et axiomes de justice tels qu'ils sont discutés dans la théorie. Dans la deuxième approche, quasi-inductive, qui nous intéresse ici,

les expériences servent à découvrir les mécanismes qui génèrent des comportements pro-sociaux ou moraux. Aussi nous intéressons-nous ici à une littérature qui s'est basée sur une méthodologie différente, qui consiste à impliquer les sujets dans des jeux d'interaction, et à observer leur comportement. L'hypothèse étant que la morale émerge de l'interaction sociale. Les protocoles de ces expériences sont basés sur les incitations économiques (*incentivized*), c'est-à-dire que les sujets sont payés en fonction de leurs choix et du choix des autres. L'analyse consiste alors à déduire des comportements observés la place et le rôle d'un sentiment moral ou de justice dans la prise de décision des sujets. Les résultats de ces expériences, nous le verrons, avaient prétention à modifier la théorie du comportement en économie.

— o —

Dans les années 80, l'économie expérimentale commence à produire des faits stylisés qui ne peuvent être expliqués par le modèle dominant. Les expériences révélatrices de ces écarts se basent sur des jeux simples qui mettent en relations deux individus, jeu d'ultimatum, jeu du dictateur, jeu d'investissement mais aussi, un peu plus tard, sur des jeux impliquant des groupes d'individus plus ou moins grands comme le jeu de contribution à un bien public (voir Camerer et al. 2003 pour une revue de cette littérature). Dans le jeu de l'ultimatum une première personne (joueur A) doit décider de la façon de partager une certaine somme d'argent avec une seconde personne, le récipiendaire, (joueur B). Dans une deuxième étape, celui-ci doit alors décider s'il accepte ou refuse l'offre. En cas de refus, aucun des deux individus ne reçoit d'argent. Selon le modèle dominant, le récipiendaire devrait accepter toute proposition qui lui laisse un paiement strictement supérieur à zéro et le sujet qui propose devrait, en conséquence, lui laisser le plus petit montant possible. Or, les expériences montrent que le récipiendaire refuse avec une fréquence élevée toute proposition inférieure à 25% du montant total et les partages proposés proches du 50/50 apparaissent avec une fréquence relativement élevée (10%). Le jeu du dictateur diffère du jeu d'ultimatum par le fait que le récipiendaire n'a pas la possibilité de refuser le partage qui lui est proposé. Les aspects stratégiques se trouvant de ce fait éliminés. Si la fréquence des propositions égoïstes augmente alors, l'écart par rapport aux comportements supposés en théorie persiste. Lorsque le protocole prend soin d'assurer l'anonymat parfait entre sujets d'une part et entre chaque sujet et les personnes organisant l'expérience d'autre part, cet écart persiste toujours (Hoffman et al. 1996).

Ces observations ont fait l'objet de deux types d'interprétation qui amenaient à reconsidérer le modèle dominant. Selon la première interprétation les sujets ont une rationalité bornée, selon la deuxième ils sont altruistes. Nous nous intéressons ici à ce deuxième type d'interprétation, de plus en plus mobilisé au détriment du premier, et aux développements théoriques qu'il a générés, en premier lieu desquels se situe la théorie des préférences sociales. Fehr et Fischbacher (2004) définissent ainsi les préférences sociales : "*other-regarding preferences in the sense that individuals who exhibit them behave as if they value the payoff of relevant reference agents positively or negatively.*" Fehr et Schmidt (1999) et Bolton et Ockenfelds (2000) proposent en effet des modèles qui, sans bouleverser le cadre dominant, permettent de reproduire ces comportements observés. Il s'agit simplement de modifier la fonction d'utilité en y intégrant une composante qui représente l'intérêt que l'individu porte au paiement des autres. Dans le modèle de Fehr et Schmidt, dit modèle d'aversion aux inégalités, cette composante est simplement la différence entre le paiement de l'individu et celui de l'autre ou une moyenne de ceux des autres. La fonction d'utilité est alors une somme pondérée du paiement matériel du joueur et de cette composante. Petite subtilité : le joueur est plus sensible aux inégalités qui sont en sa défaveur qu'à celle qui sont en sa faveur : le joueur valorise l'égalité des paiements mais il est aussi plus ou moins légèrement biaisé en sa propre faveur. Ce modèle a connu un succès phénoménal dans la littérature économique. Parmi les articles publiés dans le *Quarterly Journal of Economics* depuis sa création, c'est le quatrième article le plus cité.

L'aversion aux inégalités n'est cependant qu'une forme particulière de préférences sociales. D'autres modèles ont été proposés, notamment une préférence pour le bien-être social et une préférence pour le bien-être des plus démunis. Des expériences ont été conçues pour tenter de distinguer l'importance relative de ces différentes formes de préférences sociales (Charness and Rabin 2002, Engelman and Strobel 2004). La réciprocité est aussi considérée comme un type de préférence sociale et a occupé une place particulièrement importante dans le débat. Si, comme nous le disions plus haut, Fehr et Fishbacher (2004) donnent une définition générale des préférences sociales, ils ajoutent que la *réciprocité* est un cas particulier, plus sophistiqué que l'aversion aux inégalités : "*Strong reciprocity means that individuals behave as if their positive or negative valuation of the reference agent's payoff depends on the actions of the reference agent.*" La réciprocité a été proposée pour interpréter les résultats d'une littérature expérimentale qui semblait mettre en évidence le fait que la générosité d'un sujet dépendait de ce qu'il avait pu observer dans le passé du comportement d'un autre sujet (voir par exemple,

parmi les références les plus récentes Ben-Ner et al. 2004, Servatka 2009, Stanka 2009, Herne 2013). Plus précisément, ces travaux ont mis en évidence le fait que la générosité du dictateur pouvait être fortement corrélée à la générosité dont il avait pu bénéficier auparavant. Il y aurait au moins trois notions différentes de réciprocité que l'on peut définir facilement sur la base d'un jeu du dictateur. La *réciprocité stricte* que l'on observe dans une situation où un individu B qui a été récipiendaire d'un dictateur A et a donc subi les conséquences du choix de partage de A, se retrouve à présent dictateur pour ce même individu A. Les rôles sont simplement inversés. Dans ce type de protocole le comportement des sujets de types B s'interprète comme une expression de la réciprocité stricte lorsque il y a forte corrélation entre les choix de B envers A et les choix de A envers B. Le sujet B est généreux si le sujet A l'a été et inversement le sujet B est égoïste et donne peu ou pas si le sujet A a été égoïste. On peut observer une *réciprocité généralisée* dans un protocole où le récipiendaire B qui a observé le choix du dictateur A à son endroit devient ensuite le dictateur d'un troisième sujet C. Il y a alors réciprocité généralisée lorsque, comme précédemment, les choix du sujet B sont fortement corrélés aux choix du sujet A, indépendamment du fait que le sujet C n'est en rien responsable des choix du sujet A. Enfin, on peut observer une *réciprocité indirecte* dans un protocole où un sujet C observe le comportement d'un dictateur A à l'encontre d'un sujet B et devient ensuite dictateur pour le sujet A. La littérature montre en général l'existence de la réciprocité, surtout stricte et indirecte. Les résultats concernant la réciprocité générale sont beaucoup moins concluants.

Les différents modèles qui ont été proposés pour permettre à l'*homo economicus* de faire preuve de réciprocité ne peuvent se limiter à introduire dans la fonction d'utilité d'un individu une composante qui dépend du comportement de son partenaire. La modélisation nécessite cette fois-ci de modifier le concept d'équilibre (Falk et Fischbacher 2005). Dans les versions les plus sophistiquées proposées par Matthew Rabin (première version dans Rabin 1996) l'individu ne se préoccupe pas seulement alors de ce qu'il observe du comportement de l'autre mais aussi de ce qu'il peut en induire au sujet de ses intentions. La modélisation porte alors sur les croyances. Ces modèles sont cependant très compliqués et difficilement utilisables et résolubles. Ce qui fait dire à Ernst Fehr que son modèle d'aversion aux inégalités est une bonne approximation de ces modèles tout en étant nettement plus simple et facile à utiliser (Fehr et Fischbacher 2005, p. 152). Pourtant, si la modélisation de la réciprocité est si complexe, c'est peut-être parce qu'elle présente une différence essentielle avec l'aversion aux inégalités. La réciprocité peut difficilement se réduire à une composante de la fonction

d'utilité qui serait donc une caractéristique intrinsèque de l'individu. La modélisation de la réciprocité exige en effet de tenir compte de l'interaction entre individus. La réciprocité n'est pas une caractéristique de l'individu, elle est un mécanisme. Notons pour conclure sur ce point qu'il y a une progression dans la profondeur de l'analyse entre ces différentes modélisations lorsque l'on passe de l'idée que les sujets se préoccupent du paiement des autres à l'idée qu'ils se préoccupent du comportement des autres et enfin à l'idée qu'ils se préoccupent des intentions des autres. Les paiements pouvant résulter des comportements qui peuvent résulter eux-mêmes des intentions. Se pose-t-on la question de savoir de quoi résultent les intentions ?

---o---

Alors que l'on vient à peine de trouver ce qui est présenté comme une réponse théorique aux faits stylisés expérimentaux, le modèle des préférences sociales, Cherry et al (2002) publient de nouveaux résultats qui les remettent en question. Au protocole traditionnel basé sur un jeu du dictateur, ils proposent d'ajouter une première étape pendant laquelle le dictateur doit accomplir une tâche pour gagner le montant qu'il pourra partager dans la deuxième étape. Il apparaît que le dictateur devient alors très égoïste et son comportement devient à nouveau conforme à ce que prédit la théorie toujours dominante. Pour autant cet article ne donne pas raison à cette théorie mais montre que l'on est allé sans doute un peu vite en donnant une explication basée sur l'aversion aux inégalités. En effet, on comprend alors que les protocoles qui avaient été à la base d'une vaste littérature portaient tous d'une hypothèse fautive : les sujets n'accordent pas d'importance à l'origine du montant qu'ils ont à partager ou investir. Dans les protocoles, comme dans la fonction d'utilité, l'argent n'avait pas d'odeur. Alors que cette hypothèse devait être remise en question en même temps que l'hypothèse d'un comportement parfaitement égoïste et égocentré.

La nouvelle littérature expérimentale qui se développe alors est basée sur un protocole type composé de deux étapes : une première étape dans laquelle, à la différence du cadre de Cherry et al., tous les sujets participent et qui leur permet de gagner une somme qui peut être ensuite partagée dans la deuxième étape. On comprend que l'aversion aux inégalités a pu être une interprétation pertinente parce que l'absence de phase de production équivaut à donner des droits égaux aux deux sujets d'un jeu d'ultimatum ou du dictateur. Le partage égalitaire est alors le principe de justice le mieux adapté à ce contexte. Une autre interprétation plus générale est ainsi proposée, selon laquelle les sujets se réfèrent à des principes de justice

(Cappelen et al. 2004, Rodriguez et Moreno 2012) ou des droits, *claims* en anglais (Gächter et Riedl 2005 et 2006). Ainsi l'article de Cherry, Frykblom et Shogren a ouvert la voie à une littérature sur les principes de justice.

Dans le protocole proposé par Cappelen et al (2007), la phase de distribution, qui est un jeu du dictateur, est précédée d'une phase de production pendant laquelle les sujets doivent répondre à un questionnaire de culture générale. Dans chaque paire, le montant qui va être partagé dépend du nombre de réponses correctes mais aussi d'un prix qui n'est pas le même pour les deux sujets. Ainsi, trois points focaux peuvent apparaître qui dépendent de trois principes de justice fondamentaux, issus d'un long débat sur la responsabilité (voir Konow 2003). Le partage strictement égalitaire du montant total, le partage libertarien qui attribue à chaque sujet un montant proportionnel à ce que sa participation a permis d'obtenir, indépendamment du fait qu'il n'est pas responsable du prix qui s'applique à ses réponses correctes, et enfin le partage libertarien-égalitaire en proportion du nombre de réponses correctes. Cappelen et al. montrent qu'effectivement, les partages proposés sont très proches des partages qui correspondent aux trois principes de justice. Ils proposent alors un modèle qui leur permet d'assez bien répliquer leurs observations. Le modèle intègre dans la fonction d'utilité une composante représentant la valorisation des principes de justice. Ils remplacent l'écart entre le paiement matériel du joueur et la moyenne des paiements des autres qui représente l'aversion aux inégalités dans le modèle de Fehr et Schmidt (1999), par un écart entre le paiement matériel du joueur et un partage qui correspond à un idéal de justice. Cet idéal de justice est propre à l'individu, il le caractérise, comme les poids qui là aussi pondèrent les différentes composantes de la fonction d'utilité. Ainsi passe-t-on de la théorie de l'aversion aux inégalités à la théorie des idéaux de justice.

Plus récemment, en 2013, Rodriguez et Moreno ont proposé une conception alternative de la façon dont les principes de justice comptent dans les choix des sujets. Leur protocole est assez proche de celui de Cappelen et al. (2004) et comme ces derniers, ils montrent que le dictateur n'est pas égoïste et semble se référer à l'un ou l'autre des trois principes de justice. Cependant, en comparant un traitement où le dictateur bénéficie d'une rémunération plus élevée que le récipiendaire et un traitement avec les attributions inverses, ils montrent que les propositions des dictateurs du premier traitement sont plus fréquemment proches du principe libertarien alors que les propositions des dictateurs du deuxième traitement sont plus fréquemment proches du principe égalitaire. La différence entre les deux traitements est significative. Le traitement statistique semble donc montrer que le dictateur *choisit* le principe

de justice qui maximise son paiement. Rodriguez et Moreno montrent ainsi qu'un phénomène bien connu par ailleurs, le biais égoïste, s'applique aussi aux principes de justice. Selon Babcoq et Lowenstein (1997) le biais égoïste « *is the tendency for parties to arrive at judgments that reflect a self-serving bias to conflate what is fair with what benefits oneself.* » Ce mécanisme a été largement analysé dans la littérature en psychologie qui montre qu'il ne peut être considéré comme rationnel au sens où il n'est ni stratégique ni même conscient. Au-delà de ce mécanisme particulier, les travaux de Rodriguez et Moreno montrent surtout qu'il est possible d'avoir une autre représentation de la façon dont les principes de justice influencent le comportement des individus. Si le dictateur choisit le principe qui maximise son paiement matériel, alors cela signifie que l'on peut difficilement considérer qu'il est caractérisé par un idéal de justice comme défini dans le modèle de Cappelen et al.

Aux interprétations précédentes en termes de principes de justice s'ajoutent les interprétations en termes de droit. Dans les protocoles de Gächter et Riedl (2005 et 2006) les sujets gagnent, comme précédemment, ce qui doit être partagé mais par ailleurs le partage est dans un premier temps imposé par l'expérimentateur. Puis, de façon aléatoire, le montant à partager est réduit et les sujets qui ne peuvent plus obtenir autant qu'avant doivent alors négocier une répartition du montant réduit. Cette fois-ci la négociation peut comporter plusieurs offres et contre-offres. Gächter et Riedl montrent que le premier partage imposé par l'expérimentateur à la première étape constitue une référence tout au long de la négociation. Les sujets l'utilisent comme une revendication (*claim*). Ainsi, les sujets considèrent qu'ils ont des droits sur le montant à partager non seulement parce qu'ils l'ont gagné et *mérité* mais tout aussi bien parce que l'expérimentateur ou le hasard le leur a attribué. A la lumière de cette nouvelle interprétation, on peut reconsidérer les résultats des expériences basées sur le jeu du dictateur décrites plus haut. Le comportement égoïste du dictateur qui, ne l'oublions pas, est aussi observé, peut être alors interprété comme l'exercice d'un *droit* attribué par l'expérimentateur.

Conclusion de cette section

Que peut-on conclure de ces expériences et des tentatives d'interprétation et de modélisation qui en ont découlé ? Les modèles de Fehr et Schmidt (1999) et Cappelen et al. (2007), proposés pour interpréter l'observation de comportements pro-sociaux sont fondamentalement très proches. Ils partagent en particulier cette hypothèse selon laquelle les préférences de l'individu sont stables, conformément au modèle dominant. Dans le premier

modèle l'individu est caractérisé par son degré d'aversion aux inégalités mesuré par les paramètres que constituent les coefficients de pondération de sa fonction d'utilité. Dans le deuxième modèle l'individu est caractérisé à la fois par les coefficients de pondération et par son idéal de justice. Dans les deux cas, les paramètres sont autant de degrés de liberté qui rendent peu convaincants les résultats montrant que le modèle permet de reproduire les observations. Cependant, l'article de Ferh et Schmidt (1999) comportait une autre dimension qui a été largement occultée dans l'abondante littérature qui y a fait référence, y compris par les auteurs de l'article eux-mêmes. L'article montrait en effet que la fonction d'utilité incorporant de l'aversion aux inégalités pouvait générer des comportements opposés, coopératifs ou compétitifs, selon le type de jeu, d'ultimatum ou de marché, dans lequel les sujets étaient impliqués. Ferh et Schmidt proposaient ainsi de modéliser des préférences stables mais dont « l'expression » en termes de comportement, dépendait de l'environnement, du type d'interaction sociale. Le peu d'écho de ce résultat dans la littérature montre à quel point les sciences économiques ne sont pas prêtes à s'intéresser à l'effet de l'environnement sur le comportement des individus.

La critique essentielle que nous porterons à cette approche est justement qu'elle reste dans le cadre de l'individualisme méthodologique et conduit à une impasse. Ce cadre ne peut en effet constituer une prémisse à une théorie positive de la morale. Comment, en effet, lorsque l'explication de la morale est à chercher dans les caractéristiques de l'individu, peut-on comprendre l'émergence de la morale, c'est-à-dire comment les individus acquièrent des préférences sociales ? La seule solution pour sortir de cette contradiction a été le recours aux jeux évolutionnaires. Dans ce cadre, ce ne sont pas les individus qui apprennent mais la société. Les comportements des individus intrinsèquement moraux leurs assurent une plus grande descendance. Ainsi les comportements moraux, pro-sociaux, se trouvent-ils « sélectionnés ». Cependant, ce qu'il y a de remarquable dans la morale, c'est que dans toute société existent tout à la fois des comportements moraux et immoraux, la plus grande bonté et la plus grande cruauté. De plus, à l'échelle de l'histoire humaine on n'est pas sûr que l'une l'emporte sur l'autre et l'on ferait preuve d'un optimisme naïf en croyant qu'il existe toujours plus de moralité et que les sociétés évoluent de façon monotone en la matière. Comment alors expliquer non seulement les comportements moraux mais aussi les comportements immoraux ? Les comportements moraux ne seraient jamais complètement éliminés car ils réapparaîtraient en mutations ? Qu'est-ce qui justifierait alors d'expliquer les comportements

pro-sociaux par des mécanismes de sélection et les comportements asociaux par des mutations ?

Les travaux de Rodriguez et Moreno d'une part et de Gachter et Riedl d'autre part, sont fondamentalement différents des travaux précédents. Tout d'abord, ils laissent imaginer que les *droits* agissent comme des contraintes plutôt que comme des préférences. On pourrait donc tout aussi bien explorer une modélisation, là encore peu coûteuse, non pas des préférences sociales mais des *contraintes sociales*. Il s'agirait d'intégrer droits et principes de justice non pas dans la fonction d'utilité mais comme une contrainte, au même titre que la contrainte budgétaire. Encore faudrait-il pouvoir déterminer l'origine de cette contrainte. Par exemple, le fonctionnement de l'individu en société lui permettrait d'acquérir un capital de droits et de principes de justice qui, à l'opposé du budget, n'augmenterait pas ses possibilités mais les diminuerait. Nous n'irons pas plus loin ici dans cette direction et il y a une bonne raison pour laquelle Rodriguez et Moreno d'une part et de Gachter et Riedl d'autre part ne l'ont pas explorée. La raison en est certainement que ces auteurs, reconnaissant que les interactions et l'environnement sont déterminants pour la moralité de l'individu, ne pouvaient plus emprunter le chemin de l'individualisme méthodologique. Il y aurait eu une contradiction fondamentale entre l'observation et l'interprétation d'une part et la modélisation d'autre part.

En effet, la conclusion que l'on peut tirer de cette courte revue de la littérature expérimentale sur les préférences sociales et la morale est qu'il y a débat entre deux interprétations des comportements pro-sociaux. D'un côté une interprétation selon laquelle ces comportements sont la manifestation de caractéristiques intrinsèques aux individus et qui mène à une modélisation dans le cadre de l'individualisme méthodologique. De l'autre une interprétation selon laquelle ces comportements sont le résultat de mécanismes d'interaction entre les individus dans un environnement donné qui ne peut mener au même type de modélisation et nous demandera davantage d'inventivité.

Il n'y a pas de contradiction fondamentale entre les deux interprétations car s'intéresser aux mécanismes ne signifie pas que l'on considère que les individus n'ont pas d'idéaux de justice. Cependant, lorsque l'on s'intéresse aux mécanismes on ne peut éluder le fait que les individus ont *acquis* ces idéaux et que ce dernier aspect du problème est sans doute ce qu'il y a de plus important à comprendre. Il y a donc par contre une contradiction fondamentale entre les outils de modélisation auxquels conduisent l'une et l'autre de ces interprétations.

2. *Interactions sociales et sentiments moraux chez Adam Smith*

La question des mécanismes par lesquels les individus acquièrent des idéaux et une moralité n'est pas un impensé de la science économique. Loin de là, elle est même peut-être une de ses questions fondamentales puisque c'est finalement celle posée par Adam Smith dans la *Théorie des sentiments moraux*. Que fait Adam Smith dans cet ouvrage ? Il analyse la façon dont les individus acquièrent des sentiments moraux grâce à un apprentissage informel qui passe par le vecteur des interactions sociales. En reformulant cette question pour utiliser des termes contemporains, nous dirons qu'il s'agit de comprendre les mécanismes psychologiques qui façonnent le comportement de l'individu, lorsque celui-ci est impliqué dans des interactions sociales. Quels sont les éléments de réponse proposés par Smith ? Selon Ashraf, Camerer et Lowenstein (2005), « *Smith argued that behavior was determined by the struggle between what [he] termed the “passions” and the “impartial dictator”.* » Parmi les passions de l'individu, Smith range la faim, la douleur, la colère, le désir sexuel mais aussi et même surtout ce qu'il appelle la *sympathie*, qui permet à l'individu de ressentir les émotions de l'autre ; celle de la mère envers son enfant, celle de l'individu qui ressent la souffrance d'un autre. Cependant, chez Adam Smith, cette sympathie « passionnelle » qui serait plutôt aujourd'hui appelée empathie compassionnelle, n'est pas la garantie d'un comportement moral ; l'individu sous-estimant ou surestimant les sentiments de l'autre. Chez Smith l'empathie a donc la même place et fonction que les autres passions auxquelles l'individu est soumis, faim, peur, ou douleur et que l'on peut aussi interpréter comme des instincts égocentrés. Selon Smith l'individu peut cependant contrôler ces passions grâce au fait qu'importe aussi beaucoup pour lui le jugement des autres. L'individu voit son propre comportement du point de vue des autres, mécanisme personnalisé chez Smith par le *spectateur impartial* « looking over the shoulder of the economic man, scrutinizes every move he makes » (Grampp, 1948, p. 317). Ainsi, l'empathie comme « passion », proche de ce que les psychologues contemporains appellent l'empathie émotionnelle, n'est pas pour Smith d'une grande utilité sociale car c'est bien le spectateur impartial et donc le souci du regard de l'autre qui garantit le comportement moral. On pourrait être tenté de dire que chez Smith l'individu acquiert des sentiments moraux parce que son égocentrisme est contrebalancé par la préoccupation du jugement des autres, ce qui correspondrait à l'interprétation de Camerer et Lowenstein (2005). Pourtant les choses ne sont pas si simples. En effet, le spectateur impartial ne parvient à civiliser l'individu que parce que celui-ci a aussi un sentiment naturel de ce qui est équitable (*fairness*). Le sentiment de justice fait ainsi partie des passions naturelles de l'individu et constitue pour Smith le ciment de la société humaine. Enfin, le

spectateur impartial ne pourrait fonctionner sans la sympathie, qui fait aussi partie des passions.

La particularité remarquable de l'analyse de Smith par rapport à certaines des analyses évoquées précédemment est qu'elle porte à la fois sur l'individu, qu'il étudie en profondeur dans son comportement et son fonctionnement et sur les relations entre cet individu et la société dans son ensemble. Smith ne se contente pas de montrer que l'individu a ce qui pourrait entrer dans la catégorie de ce que l'on appelle aujourd'hui des préférences sociales, il cherche à décrypter les mécanismes qui expliquent pourquoi et comment. Or, ce faisant, Smith montre que la pensée, les sentiments moraux, le comportement de l'individu sont façonnés par l'interaction sociale. En cela sa démarche ne sacrifie en rien à l'individualisme méthodologique.

Quand Adam Smith écrit la *Théorie des sentiments moraux* il s'intéresse à des relations entre individus de nature très différentes. Il semble mettre l'accent sur les relations directes, les protagonistes pouvant s'écouter et s'observer, mais il fait aussi référence à des relations plus indirectes. Il s'intéresse par exemple à des relations entre groupes sociaux, comme les pauvres et les puissants. Les uns ayant des sentiments moraux envers les autres et inversement. Son analyse se veut donc très générale. Par contre, lorsque il écrit *la Richesse des nations*, publiée dix-sept ans après la première édition des *sentiments moraux*, il s'intéresse à une relation particulière, la relation marchande. Le concept de sympathie a alors complètement disparu. Une peut-être trop célèbre citation de Smith illustre bien la différence entre les deux:

« *Ce n'est pas de la bienveillance du boucher, du marchand de bière et du boulanger, que nous attendons notre dîner, mais du soin qu'ils apportent à leurs intérêts. Nous ne nous adressons pas à leur humanité, mais à leur égoïsme ; et ce n'est jamais de nos besoins que nous leurs parlons, c'est toujours de leur avantage.* »

Smith dit clairement que ce type de relation d'interdépendance égoïste est à la base de la division du travail (livre premier, deuxième chapitre) et donc de *la Richesse des nations*, « *cette opulence générale qui se répand jusque dans les dernières classes du peuple.* » La bienveillance désigne par contre la relation à l'autre qui peut être basée sur l'empathie (ou la sympathie Smithienne). Si l'on prend l'œuvre de Smith dans son ensemble, on doit donc distinguer deux types de relations qui font l'individu en société. D'une part, la sympathie et le spectateur impartial qui le poussent à acquérir des sentiments moraux dont la bienveillance, et d'autre part, l'intérêt personnel qu'il a à tirer d'une interdépendance économique. L'individu

est ainsi doublement lié à la société. Rien n'oblige à penser pour autant que Smith ait été schizophrène, nous considérons plutôt qu'il a conçu ces deux analyses en complémentarité l'une de l'autre. Ainsi dit-il dans *la Richesse des nations* que « *l'homme a presque continuellement besoin du secours de ses semblables, et c'est en vain qu'il l'attendrait de leur seule bienveillance.* » Cette phrase de Smith montre que la dépendance de l'individu vis-à-vis des autres est telle, que la bienveillance des autres n'y suffirait pas mais elle ne signifie pas pour autant que la bienveillance n'est pas aussi nécessaire.

Cependant, cette apparente opposition entre le Smith de la *Théorie des sentiments moraux* et celui de *la Richesse des nations* ne permet pas, selon nous, de justifier le désintéret des économistes pour la morale. Tout d'abord, quand bien même Smith aurait considéré que les activités marchandes pouvaient se passer de bienveillance de la part de ses participants, ses conclusions seraient peut-être différentes aujourd'hui, alors que des activités qui ne sont pas a priori purement marchande le deviennent ou se trouvent soumises à ce que l'on appelle une « logique économique ». La pollution est un droit échangeable, les hôpitaux sont soumis à des contraintes budgétaires et les médecins se plaignent des interférences de ces logiques dans leur pratique. L'extension de la sphère marchande justifie aujourd'hui que vertus et bienveillance persistent dans ces activités.

Le Smith de *la Richesse des nations* fait référence à l'égoïsme dans l'échange. Mais qu'en est-il de la production ? Dans la mesure où, étant donnée la division du travail, l'individu produit pour échanger, on a pu considérer que l'égoïsme de *la Richesse des nations* concernait toutes les relations économiques. La critique des philosophes des vertus et de MacIntyre en particulier lève une ambiguïté à ce sujet. Dans *Après la vertu* (1981), MacIntyre, illustre son propos selon lequel « *nos obligations morales proviennent des pratiques sociales de la communauté à laquelle nous appartenons* », par une série d'exemples dans lesquels les pratiques sociales sont des pratiques liées à des activités de production. Il en déduit que l'échange présente le risque de corrompre les vertus associées à ces activités de production car le marché est purement instrumental. Autrement dit, l'échange n'a pas de fin en soi, n'a pas de *telos*, son unique intérêt est d'apporter à celui qui y prend part des biens externes à cette pratique, qui lui procurent une satisfaction purement égoïste comme une rémunération.

L'économie comportementale, inspirée aussi de la psychologie sociale, a repris la critique de MacIntyre. Un premier résultat de cette littérature a été de montrer que les individus peuvent avoir deux types de motivations lorsqu'ils participent à une activité, des motivations externes

et internes, à rapprocher des biens externes et internes de MacIntyre. Les motivations externes sont les motivations qui ne sont pas spécifiques à l'activité concernée et sont purement égoïstes, comme une rémunération monétaire. Les motivations internes sont plus complexes, sociales et morales. Par exemple, un individu qui décide de donner son sang poursuit des motivations internes, c'est-à-dire soit morales, pour aider, soit de positionnement social, pour le dire simplement, pour être bien vu. Un deuxième résultat important de cette littérature, conséquence de cette distinction entre motivations internes et externes, est de montrer que ces dernières peuvent chasser les premières. Ce phénomène est appelé le *crowding out*. L'exemple le plus célèbre est celui de la collecte de sang décrit par Richard Titmuss qui montre, en comparant les systèmes de collecte de sang aux Etats Unis et en Grande Bretagne, qu'un système basé sur le don s'avère plus efficace qu'un système avec rémunération. Instaurer une rémunération pour accroître la collecte de sang peut avoir l'effet inverse. On comprend alors que l'on ne peut plus simplement admettre comme point de départ que les individus ont des sentiments moraux dans certaines activités et sont purement égoïstes dans les relations marchandes ou économiques, on doit considérer que c'est un résultat. Un économiste libéral peut objecter que, lorsque l'on étudie les activités marchandes, il revient au même de partir de l'hypothèse que l'agent économique n'a pas de sentiments moraux ou de concevoir que l'individu puisse être plus sophistiqué mais perd ses sentiments moraux en se livrant à ces activités. Il faut cependant avoir une foi indéfectible dans le libéralisme économique pour ne pas voir le risque que comporte une telle confusion. Si l'on admet que le fonctionnement du marché n'est pas parfait, si l'on imagine par exemple qu'une quelconque intervention de l'Etat est nécessaire, alors la potentialité de sentiments moraux ne peut être ignorée. Par ailleurs, même le plus libéral des économistes peut garder comme un idéal le fonctionnement harmonieux d'individus purement égoïstes dans le cadre du marché mais il se doit d'admettre que cet idéal est loin d'être réalisé. Peu importe alors que la faute en revienne à l'Etat ou au marché lui-même, dans cet état bien réel, qui n'est peut-être qu'intermédiaire, les sentiments moraux peuvent avoir un rôle à jouer. Enfin, l'idée des motivations internes et externes de l'économie comportementale est bien de montrer l'interférence de la sphère marchande sur les activités qui n'en font pas partie. Comment alors justifier que les économistes puissent continuer à s'intéresser aux activités marchandes sans se préoccuper de l'influence que ces activités pourraient avoir sur le reste des activités humaines ?

Si l'on accorde le minimum d'importance à ces questions, alors on admettra que l'on ne peut isoler les types de comportement en fonction des types d'activités et l'on devra conclure

qu'en économie, on ne peut se baser sur une théorie du comportement incomplète, en particulier une théorie qui ne tient pas compte des sentiments moraux.

3. *L'apport des neurosciences sociales*

L'hypothèse empathie-altruisme

Puisqu'il n'est pas possible de construire une théorie du comportement qui puisse faire abstraction d'une théorie positive de la morale et qui soit spécifique aux activités économiques, il est légitime d'aller voir comment les autres sciences du comportement pensent ces questions. On est alors frappé de voir que l'empathie, comme chez Smith, y occupe une place importante. Pour l'éthologue Frans de Waal (2010), c'est l'empathie qui est à la base de la capacité à coopérer chez l'être humain, comme chez tous les mammifères. Dans les neurosciences, les recherches sur l'empathie se sont accélérées depuis la deuxième partie des années 1990, à la suite de la découverte par l'équipe de chercheurs en neurosciences de Giacomo Rizzolatti (Rizzolatti et al. 1996) des neurones miroirs, aussi appelés neurones empathiques. Ces neurones du cerveau ont la particularité de présenter une activité aussi bien lorsqu'un individu exécute une action que lorsqu'il observe un autre individu exécuter la même action. Cette découverte semblait déterminante car elle conforte une approche naturaliste de l'empathie. Et si l'empathie pouvait être décrite par le fonctionnement d'une catégorie de neurones ?! Mais la fonction des neurones miroirs, dont l'existence chez l'humain n'a que récemment été prouvée (Mukamel et al. 2010), reste un domaine de recherche balbutiant ; et l'empathie reste un mécanisme multiforme et complexe.

Psychologues, chercheurs en neurosciences et économistes observent les comportements des sujets dans des expériences de laboratoire basées sur des protocoles parfois identiques, parfois marqués des spécificités de chaque discipline. Jeu d'ultimatum, jeu du dictateur et jeu de la confiance (ou jeu d'investissement) sont les principaux ingrédients de ces protocoles. Ces différentes disciplines observent toutes, de la part des individus, des comportements dits *pro-sociaux*, pour le dire vite, bénéfiques aux autres. Cependant, ces comportements amènent les uns et les autres à des interprétations différentes. Selon bon nombre de psychologues et chercheurs en neurosciences l'empathie dans ses différentes déclinaisons est à l'origine de ces comportements. C'est ce que le psychologue social Daniel Batson (1991) appelle l'hypothèse Empathie-altruisme. Pour les économistes ce sont les préférences sociales qui sont à l'origine des comportements pro-sociaux. Existe-t-il un pont entre ces deux interprétations ? Tania

Singer, chercheuse en neurosciences, et Ernst Fehr (Singer and Fehr 2005) considèrent que l'empathie permet de justifier le modèle des préférences sociales. Ils encouragent, dans leur article, au développement d'un programme pour le nouveau domaine de la neuro-économie, dont l'objectif serait de montrer que les mécanismes d'empathie sont bien à l'origine des préférences sociales. Ils pointent en particulier le fait qu'il y a la même hétérogénéité en termes d'empathie mesurée par des tests appropriés, que celle supposée dans le modèle des préférences sociales. Mais ce faisant, ils semblent établir une simple identité entre les deux concepts et font passer au second plan la question des mécanismes par lesquels l'empathie pourrait générer des préférences sociales. Les préférences sociales, sous-entendu sous la forme de l'aversion aux inégalités, seraient le pendant dans le langage économique de l'empathie dans le langage des neurosciences. Comme si l'empathie était la traduction dans le cerveau et donc la justification physiologique du modèle des préférences sociales. Il est intéressant de remarquer que l'individualisme méthodologique de la théorie des préférences sociales amène facilement au naturalisme. Or, cette identification des deux concepts ne va pas de soi. En particulier, comme le disent Kirman et Teschl (2009), dans le modèle de préférences sociales de Fehr et Schmidt « ... one is not concerned with the situation of another individual but the state of the population with respect to oneself. » Or, peut-il y avoir une relation directe entre l'empathie qui concerne la relation entre deux individus et les préférences sociales qui représentent la relation entre l'individu et la société dans son ensemble ? Telle est bien la question traitée par Adam Smith dans la *Théorie des sentiments moraux* et la complexité de la réponse qu'il y apporte montre toute l'attention qu'il faut apporter à cette relation.

Composante émotionnelle et théorie de l'esprit

Les sciences du comportement ne se limitent pas à l'hypothèse d'une relation entre empathie et altruisme. Leur mission est d'expliquer la nature exacte de ce lien, d'en décrypter les mécanismes. Une condition préalable a été de catégoriser les différentes composantes de l'empathie. Frans de Waal donne une définition englobante de l'empathie. C'est la capacité d'un individu à s'identifier à ceux qui sont dans le besoin et la douleur ; cette identification suscitant une émotion qui pousse l'individu à tenter d'aider. La psychologie et les neurosciences tentent au contraire de définir avec précision les différentes composantes de ce mécanisme qu'elles décrivent comme complexe. Au moins deux composantes sont distinguées. La composante émotionnelle désigne la capacité d'un individu à partager les émotions d'un autre alors que la composante cognitive ou théorie de l'esprit désigne sa

capacité à se représenter les intentions et croyances de l'autre. La composante émotionnelle de l'empathie semble être déclenchée par des images ou des représentations, en tout cas la possibilité d'identifier l'objet de l'empathie. Les économistes Small, Loewenstein et Slovic (2007) montrent par exemple que les sujets sont beaucoup plus généreux envers des victimes qu'ils ont pu identifier. Les conditions de mise en œuvre de la théorie de l'esprit sont moins claires. Selon la théorie de la simulation mentale, l'individu opère une projection qu'il corrige pour établir une différence entre lui et l'autre, objet de son empathie. Cette correction se fait grâce à une série de critères d'ajustement tirés de la prise en compte de l'environnement social, historique, culturel de la personne-cible (Davies 1995). Par ailleurs, les psychologues Decety et Cowell (2014) distinguent une troisième composante, motivationnelle, qui désigne la volonté d'un individu de prendre soin d'un autre et qui n'est pas la conséquence nécessaire de l'une ou l'autre des deux autres composantes. Selon ces chercheurs, ces différentes composantes de l'empathie correspondent à des mécanismes très différents qui peuvent interagir ou au contraire fonctionner en parallèle.

Cette caractérisation de l'empathie a permis de mettre en évidence la complexité de la relation entre les différentes composantes. Elle a aussi conduit à reconnaître que la relation entre ces différentes composantes de l'empathie et les différentes composantes de l'altruisme est même plus complexe. Tout récemment, les économistes Artinger, Exadaktylos, Koppel et Saaksvuori (2014) ont répondu à la proposition de Singer et Fehr et testé la relation entre l'empathie et les comportements supposés être issus des préférences sociales. Ils ont ainsi construit un protocole basé sur le jeu du dictateur et le jeu d'ultimatum et tenté de corréliser la générosité des dictateurs aux résultats de différents questionnaires permettant d'évaluer la composante émotionnelle de l'empathie (Interpersonal Reactivity Index ou IRI) et la théorie de l'esprit. Or, contrairement à l'idée reçue selon laquelle la générosité serait émotionnelle et l'égoïsme rationnel, ils ne trouvent aucune corrélation significative entre la générosité des dictateurs et leur empathie émotionnelle mais montrent par contre que les dictateurs généreux sont ceux qui ont les performances les plus élevées dans les capacités associées à la théorie de l'esprit. Leurs résultats sont donc en contradiction avec bon nombre d'expériences plus anciennes de psychologues qui semblent confirmer l'hypothèse d'une corrélation entre l'empathie conçue comme émotionnelle et altruisme. Jean Decety et Keith Yoder (2015), respectivement chercheur en neurosciences et psychologue, ont montré très récemment un résultat allant dans la même direction. Dans leur expérience ils s'intéressent au sentiment

d'injustice et montrent que ce n'est pas la composante émotionnelle de l'empathie mais sa composante cognitive qui permet de prédire l'importance de ce sentiment chez les individus.

Groupes et communauté empathique

Decety et Cowell (2014) montrent un autre aspect de la complexité de la relation entre comportements pro-sociaux et empathie : l'empathie ne conduit pas nécessairement à des comportements pro-sociaux. A l'instar de Frans de Waal ou d'Adam Smith, ils considèrent que l'empathie, définie comme la capacité d'un individu à partager les émotions d'un autre, décrit au départ les relations entre la mère et l'enfant. Elle peut alors, par apprentissage, s'étendre aux relations entre les membres d'un groupe de proches. Cependant, cette même empathie qui rapproche les membres d'un groupe peut motiver des comportements de rejet et d'agressivité vis-à-vis des individus d'un autre groupe. La littérature expérimentale en psychologie sociale mais aussi en économie, s'est beaucoup intéressée à ces comportements *in-groups* et *out-groups*. Ce mécanisme semble si fort que nombres d'expériences en économie ont pu mettre en évidence, dans le laboratoire, des différences de comportement entre des membres de « groupes » constitués de manière parfaitement exogène et sans pertinence. Dans l'expérience de Chen et Li (1999) il a suffi de désigner les membres d'un groupe par une couleur, bleue ou jaune, pour générer des comportements beaucoup plus coopératifs entre sujets d'une même couleur.

Le rôle de l'empathie dans les comportements de groupe n'est cependant pas parfaitement clair. De nombreuses questions subsistent. L'empathie semble s'exercer parmi les membres d'un groupe et les sciences cognitives expliquent comment l'individu apprend à étendre son empathie au groupe. Mais quel est le rôle de l'empathie à l'égard des individus d'un autre groupe ou simplement identifiés comme étant hors du groupe ? Doit-on parler simplement d'absence d'empathie. Les mécanismes d'empathie seraient alors bloqués lorsque l'individu ne reconnaît pas chez l'autre des caractéristiques de son groupe. Nous pouvons cependant penser que le rôle de l'empathie est plus actif. Lorsque l'individu ne reconnaît pas chez l'autre des caractéristiques de son groupe ou plutôt lorsqu'il reconnaît chez l'autre des caractéristiques d'un autre groupe, son empathie *positive* pourrait se transformer en un *sentiment négatif* et générer de l'agressivité. En effet, le comportement négatif vis-à-vis des individus extérieurs est peut-être vécu comme un comportement positif vis-à-vis des membres du groupe. Dans cette interprétation il faut considérer le lien à l'autre comme positif ou négatif. Il faudrait pouvoir alors définir l'antonyme d'empathie. Est-ce l'absence d'empathie ?

Ou plutôt une forme de cruauté ? Là aussi il serait nécessaire de tenir compte de la catégorisation des différentes composantes de l'empathie. Qu'est-ce alors que l'antonyme de l'empathie émotionnelle ? Est-ce l'absence d'émotion, la froideur, en soi déjà monstrueuse, ou la capacité à ressentir des émotions inverses ? Comme jouir de la souffrance de l'autre ? Et comment définir l'antonyme de l'empathie cognitive ? Comme une forme d'autisme, d'incapacité à comprendre l'autre ? Ou est-ce utiliser cette compréhension de l'autre à des fins stratégiques pour soi ou contre l'autre ?

Les neurosciences sociales

Jean Decety, chercheur au Social Cognitive Neuroscience Laboratory de l'Université de Chicago, est un des fondateurs des neurosciences sociales. Il défend l'idée que l'étude du cerveau doit s'accompagner d'une vision d'ensemble de la société. Il reproche en effet aux neurosciences classiques une tendance à focaliser leurs analyses sur le cerveau, en le considérant comme un objet isolé, alors que sa fonction essentielle est d'interagir avec son environnement, en particulier social. En quelque sorte, Decety fait aux neurosciences la même critique que nous adressons ici aux sciences économiques, si ce n'est qu'il parle du cerveau de l'individu quand nous parlons de la façon dont celui-ci s'en sert.

L'objectif essentiel des neurosciences sociales est donc de comprendre comment le cerveau de l'individu interagit avec son environnement social. L'empathie est bien évidemment un mécanisme clef de cette interaction. Decety l'étudie avec soin. Cependant, par rapport aux autres chercheurs en neurosciences qui travaillent sur le même thème, il a pour particularité de tenter de dépasser l'analyse de l'empathie chez l'individu. Plus exactement, il tente de changer de point de vue et pose des hypothèses sur les mécanismes sociaux qui pourraient façonner la capacité empathique de l'individu. Ainsi, dans un article récent co-écrit avec Cowell (Decety et Cowell 2014), ces auteurs posent l'hypothèse que la morale est une tentative de la société d'étendre l'empathie émotionnelle, sentiment primitif qui caractérise avant tout les relations entre la mère et l'enfant, aux relations entre tous ses membres. A la limite, l'humanisme serait une tentative d'étendre ce sentiment à l'humanité entière. Ainsi, la morale serait une institution destinée à exploiter le potentiel de cohésion de l'empathie en tentant d'éliminer son potentiel de scission que nous évoquons plus haut. Decety commence par étudier l'empathie comme un mécanisme qui, chez l'individu, le connecte à son environnement social. Puis il pose des hypothèses sur ce qui, dans cet environnement social, façonne le comportement de l'individu par le biais de ce mécanisme. Nous avons parlé de son hypothèse sur la morale mais il s'intéresse aussi à la justice et à la littérature. Mais d'où viennent ces forces sociales qui agissent sur l'individu ? Sans répondre à cette question on court le risque d'attribuer à la société et ses institutions une volonté particulière. La boucle est loin d'être bouclée. Pourra-t-on aller plus loin dans l'intégration des deux approches pour comprendre l'interaction entre individus et société dans la production de la morale ? L'élaboration d'une théorie positive de la morale est encore un projet, un défi des sciences humaines et sociales qui est loin d'être gagné.

Conclusion

Plutôt que de poursuivre le chemin qui a mené l'économie de *la Richesse des nations* aux préférences sociales, nous avons voulu montrer qu'il existe une voie différente, sur les traces de l'Adam Smith de la *Théorie des sentiments moraux*, qui pourrait bénéficier des développements récents de la psychologie et des neurosciences sociales. A l'écart du modèle de la maximisation sous contrainte, cette voie s'intéresse aux mécanismes par lesquels l'individu interagit avec son environnement social et institutionnel. Le rapport aux autres, de secondaire dans le modèle dominant, devient alors déterminant, faisant de l'empathie, comme mode de fonctionnement, et de la morale, comme institution, les clefs qui permettront peut-être d'en comprendre les ressorts.

Certes nous sommes encore loin d'une théorie positive de la morale. Jusqu'à présent, les principales avancées de la recherche dans les neurosciences ont concerné les différentes formes de l'empathie, c'est-à-dire les différents mécanismes physiologiques et cognitifs par lesquels l'individu perçoit l'autre et la façon dont il se comporte en conséquence. La neuro-économie a participé à ces avancées. Ces recherches restent très centrées sur l'individu. Le social est résumé à un autre individu avec qui il interagit ou à un ensemble indéfini, la société. En changeant de point de vue, les neurosciences sociales se sont aussi intéressées à ce qui dans l'environnement social de l'individu, façonne son comportement par le biais de ce mécanisme et surtout, façonne le mécanisme lui-même. L'interaction entre l'individu et son environnement social n'est pas cependant bouclée.

Nous nous proposons de conclure en revenant à Adam Smith pour souligner à quel point les recherches récentes n'ont fait que confirmer sa description. Les neurosciences et l'économie expérimentale ont montré que ce n'est pas l'empathie émotionnelle mais l'empathie cognitive qui serait à l'origine des comportements pro-sociaux et du sentiment d'injustice. Or Smith considérait que ce qu'il appelait la sympathie comme une passion ne suffisait pas à générer des sentiments moraux. Par contre, le mécanisme déterminant selon lui passait par ce qu'il appelait le spectateur impartial, un mécanisme cognitif complexe par lequel l'individu construit son sentiment moral en comparant son propre sentiment et celui d'un autre à partir d'un point de vue tiers qui représente celui de la société. On peut se demander si l'on ne pourrait pas davantage exploiter le modèle du spectateur impartial en le combinant notamment aux apports des neurosciences sociales pour, en particulier analyser les comportements de groupe.

Références

- Artinger F, Exadaktylos F, Koppel H, Saaksvuori L (2014), “*In Others’ Shoes: Do Individual Differences in Empathy and Theory of Mind Shape Social Preferences?*” PLoS ONE 9(4): e92844. doi:10.1371/journal.pone.0092844.
- Ashraf, Nava, Colin F. Camerer and George Loewenstein, “*Adam Smith, Behavioral Economist*”, Journal of Economic Perspectives—Volume 19, Number 3—Summer 2005—Pages 131–145.
- Babcock, Linda, and George Loewenstein. 1997. “*Explaining Bargaining Impasse: The Role of Self-Serving Biases.*” Journal of Economic Perspectives, 11(1): 109-126.
- Batson, C.D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*, Hillsdale, NJ: Erlbaum.
- Ben-Ner, Avner, Louis Putterman, Fanmin Kong, Dan Magan, “*Reciprocity in a two-part dictator game*”, Journal of Economic Behavior & Organization, Vol. 53 (2004) 333–352.
- Bolton, Gary E., and Axel Ockenfels. 2000. “*ERC: A Theory of Equity, Reciprocity, and Competition.*” American Economic Review, 90(1): 166-93.
- Camerer, C., 2003. *Behavioral Game Theory*. Princeton University Press, Princeton.
- Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen and Bertil Tungodden, “*The Pluralism of Fairness Ideals: An Experimental Approach*”, The American Economic Review, Vol. 97, No. 3 (Jun., 2007), pp. 818-827.
- Charness, Gary and Matthew Rabin, “*Understanding Social Preferences with Simple Tests*”, The Quarterly Journal of Economics, Vol. 117, No. 3 (Aug., 2002), pp. 817-869.
- Chen, Yan, and Sherry Xin Li. 2009. “*Group Identity and Social Preferences.*” American Economic Review, 99(1): 431-57.
- Cheng, Yawei, Chenyi Chen, Ching-Po Lin, Kun-Hsien Chou, Jean Decety (2010), “*Love hurts: an fMRI study.*” Neuroimage 51, 923–929.
- Cherry, Todd L., Peter Fryklblom and Jason F. Shogren, “*Hardnose the Dictator*”, The American Economic Review, Vol. 92, No. 4 (Sep., 2002), pp. 1218-1221

Davies, M. et Stone, T., (dirs.), (1995), *Mental Simulation*, Oxford, Blackwell.

Decety, Jean and Jason M. Cowell, “*The complex relation between morality and empathy*”, *Trends in Cognitive Sciences* July 2014, Vol. 18, No. 7.

Falk and Fischbacher, *Modeling Strong Reciprocity*, in *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Herbert Gintis, Samuel Bowles, Robert Boyd and Ernst Fehr Editors, 2006

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-68.

Fehr et Fischbacher, *The Economics of Strong Reciprocity*, in *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Herbert Gintis, Samuel Bowles, Robert Boyd and Ernst Fehr Editors, 2006

Gächter, Simon et Arno Riedl, *Moral Property Rights in Bargaining with Infeasible Claims*. *Management Science*, Vol. 51, No. 2 (Feb., 2005), pp. 249-263

Gächter, Simon et Arno Riedl, *Dividing justly in bargaining problems with claims. Normative judgments and actual negotiations*. *Social Choice and Welfare* (2006) 27: 571–594

Gaertner, Wulf et Erik Schokkaert. *Empirical Social Choice, questionnaire-experimental studies on distributive justice*. Cambridge University Press. 2012.

Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. 2006.

Grampp, William. 1948. Adam Smith and the Economic Man. *Journal of Political Economy*. 56:4, pp. 315-36.

Herne, Kaisa, Olli Lappalainen, Elina Kestilä-Kekkonen, *Experimental comparison of direct, general, and indirect reciprocity*, *The Journal of Socio-Economics* 45 (2013) 38– 46

Hoffman, Elizabeth; McCabe, Kevin and Smith, Vernon. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, June 1996, 86(3), pp. 653-60.

Kahneman, D., Wakker, P. & Sarin, R. (1997). “Back to Bentham? Explorations of experienced utility.” *The Quarterly Journal of Economics*, 112, 375-406.

Kirman, Alan and Miriam Teschl, “*Selfish or selfless? The role of empathy in economics*”, Phil. Trans. R. Soc. B 27 January 2010 vol. 365 no. 1538 303-317.

Kolm, S.-C. (2007), « *Macrojustice : distribution, impôts et transferts optimaux* », Revue d'économie politique, vol. 117, n°1, janv-fév. 2007, pp. 61-89.

Konow, James. “*Which Is the Fairest One of All? A Positive Analysis of Justice Theories*”, Journal of Economic Literature, Vol. 41, No. 4 (Dec., 2003), pp. 1188-1239.

Roy Mukamel, Arne D. Ekstrom, Jonas Kaplan, Marco Iacoboni, Itzhak Fried, « *Single-neuron responses in humans during execution and observation of actions* », Current Biology, Volume 20, Issue 8, p750–756, 27 April 2010.

Rizzolatti G, Fadiga L, Fogassi L, Gallese V. 1996. *Premotor cortex and the recognition of motor actions*. Cogn. Brain Res. 3:131–41.

Rizzolatti, Giacomo and Laila Craighero, *The mirror-neuron system*, Annu. Rev. Neurosci. 2004. 27:169–92

Rodriguez-Lara, Ismael and Luis Moreno-Garrido, “Self-interest and fairness: self-serving choices of justice principles”, Exp Econ (2012) 15:158–175.

Servátka, Maros. *Does generosity generate generosity? An experimental study of reputation effects in a dictator game*, The Journal of Socio-Economics 39 (2010) 11–17.

Singer, Tania, and Ernst Fehr, 2005. “*The Neuroeconomics of Mind Reading and Empathy*”, American Economic Review, 95(2): 340-345.

Small, Deborah A. & Loewenstein, George & Slovic, Paul, 2007. “*Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims*,” Organizational Behavior and Human Decision Processes, Elsevier, vol. 102(2), pages 143-153, March.

Smith Adam, 1759 (1981). *The Theory of Moral Sentiments*. D. D. Raphael and A. L. Macfie, eds. Liberty Fund: Indianapolis.

Smith Adam, 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*.

Stanca, Luca. *Measuring indirect reciprocity: Whose back do we scratch?*, Journal of Economic Psychology 30 (2009) 190–202.

Titmuss, Richard, *The Gift Relationship: From Human Blood to Social Policy* (1970).

De Waal, Frans, *The Age of Empathy: Nature's Lessons for a Kinder Society*, London: Potter Style, 2010.