

NORMES ET CHOIX RATIONNELS

Philip PETTIT

LES NORMES ET LA MAIN INVISIBLE

Qu'est-ce qu'une norme ?

Telle est la première question qui se pose. Je définis les normes par trois conditions. La première est que les normes sont des régularités auxquelles la plupart des gens, dans une population donnée, obéissent. Disons qu'une norme gouverne ce qu'on doit faire dans une situation particulière *S* : la norme dit qu'on doit se conformer à une régularité *R*. Si nous avons une vraie norme, selon ma définition – une vraie norme et pas seulement une mesure ou un modèle – alors il faut que presque tout le monde dans la société, lorsque la situation *S* se présente, conforme son comportement à *R*.

Dans une société, nous rencontrons beaucoup de régularités qu'on ne peut pas considérer comme des normes. Je pense à des régularités accidentelles ou à des régularités dont personne ne désire la réalisation. Presque tout le monde va se coucher avant minuit, mais cela n'est pas une norme. Pour qu'une régularité soit une norme, il faut qu'une deuxième condition soit remplie : presque tout le monde approuve un comportement conforme à la régularité et presque tout le monde désapprouve un comportement non conforme.

L'approbation (ou la désapprobation) est une chose complexe. Si on approuve (ou désapprouve) ce qu'un autre fait, on peut le faire pour des raisons différentes : parce que le comportement est bon pour la société, ou parce que le comportement convient à ce qu'on désire soi-même. De plus, en approuvant (ou en désapprouvant) le comportement, on peut avoir – et peut-être exprimer – des attitudes différentes : l'admiration, la gratitude ou l'acceptation, par exemple. Enfin, on peut approuver le type de comportement en général, et ce cas en particulier, ou on peut l'approuver seulement au cas par cas : *in sensu diviso*, comme le disaient les scolastiques, et non

Je voudrais examiner ici la théorie du choix rationnel en tant que théorie des normes (1). En effet, depuis quelques années, les théoriciens du choix rationnel proposent une explication des normes par la main invisible : l'existence des normes est le résultat non intentionnel de l'agrégation de décisions individuelles dans des stratégies de coordination de l'action. Après avoir présenté et critiqué cette explication des normes par la main invisible, je lui opposerai un autre mode d'explication, une explication par la main intangible. Dans les deux premières parties du texte, j'adopterai une perspective interne. Je supposerai ainsi que la théorie est valide en tant que théorie du comportement humain. Dans la dernière partie, je lèverai cette présupposition et examinerai la théorie d'un point de vue externe. Ma thèse est que la théorie du choix rationnel a une certaine validité, mais pas la validité qu'on lui a habituellement reconnue. Ma position vis-à-vis de cette théorie est œcuménique et non partisane, pluraliste et non moniste : la théorie est valide, mais en un sens qui permet aux autres théories – pas à toutes bien sûr – d'être également valides.

(1) Cet article reprend une série de conférences que Philip PETTIT a données à l'EHESS et au CREA (Ecole Polytechnique) à l'automne 1992. Louis QUÉRÉ a adapté pour Réseaux le texte présenté oralement, que Pascale GRISON avait aidé à préparer.

in sensu composito Quel est le sens pertinent pour définir les normes ? Je suppose ici que l'approbation et la désapprobation doivent être prises dans leur sens le plus large. Beaucoup de normes nécessitent l'approbation et la désapprobation en un sens étroit : il y a des normes qu'on approuve et désapprouve sur une base morale et non pour des raisons personnelles. Mais cela n'est peut-être pas vrai pour toutes les normes, et il vaut mieux éviter de faire des présuppositions contestables.

La plupart des théoriciens seront d'accord avec ce que je viens de dire sur les normes. Mais il y a une troisième condition qu'on doit ajouter aux deux premières et peu de gens ont remarqué qu'il était nécessaire de l'introduire. Imaginons qu'il y ait une régularité dans une société, que manifestement presque tout le monde connaît. Imaginons que presque tout le monde approuve le comportement conforme à la régularité et désapprouve le comportement non conforme. Mais imaginons aussi que le fait de l'approbation ou de la désapprobation n'ait aucune conséquence par rapport à ce que les gens sont disposés à faire. L'approbation et la désapprobation ne rendent pas plus probable que les gens vont se conformer à la régularité. Le fait de l'approbation et de la désapprobation n'est pas pertinent pour la conformité, il est, pourrait-on dire, épiphénoménal. Je doute que l'on puisse décrire une régularité de cette espèce comme une vraie norme, en tout cas, elle ne serait pas une norme ordinaire. En ce qui concerne les normes usuelles, et c'est là ma troisième condition, il faut que le fait que presque tout le monde approuve la conformité et désapprouve la non-conformité rende la conformité plus probable. Il faut que l'approbation et la désapprobation soient pertinentes.

Ces propositions suffisent pour définir les normes. On peut ajouter que les conditions que je viens de mentionner doivent être remplies consciemment par les acteurs : elles doivent être une affaire de *common knowledge*. Cependant je vais ignorer cette condition, parce qu'elle n'est pas pertinente pour mon propos. La défini-

tion sur laquelle sera basée la discussion est la suivante (2)

Une régularité R dans le comportement des membres d'une population P, dans une situation S, est une norme, si et seulement si

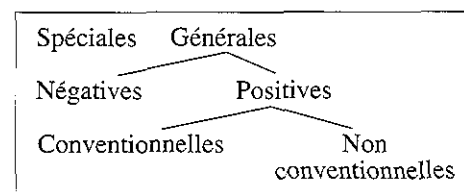
1 Presque tout le monde en P se conforme à R dans S,

2 Presque tout le monde approuve la conformité d'un autre et désapprouve sa déviation,

3 Ce fait est pertinent, entre les membres de P, pour la conformité à R, il la rend plus probable qu'autrement.

Dans une société comme la nôtre, il y a beaucoup de régularités qui sont des normes selon cette définition : dire la vérité, tenir ses promesses, ne pas voler, remplir ses devoirs civiques, contribuer aux projets collectifs. J'aurai de tels exemples à l'esprit dans la discussion qui suit. Mais je dois faire remarquer qu'il ne s'agit là que d'une espèce de normes. D'abord, elles sont toutes des normes pertinentes pour chaque adulte dans une société comme la nôtre, elles sont des normes générales, pas des normes spécifiques à des groupes particuliers. En second lieu, elles sont toutes des normes qui résolvent des problèmes collectifs et font progresser le bien-être de la société, il y a des normes qui ne résolvent pas ces problèmes et ne sont pas bonnes pour la société dans son ensemble – la norme générale de la revanche par exemple, telle qu'elle existe dans certaines sociétés. Et, en troisième lieu, elles sont des normes non conventionnelles : étant donné une convention, tout le monde ne peut pas dévier avec bénéfice, mais cela n'est pas vrai pour les normes dont nous parlons.

Nous pouvons représenter la variété des normes dans le diagramme suivant



(2) PETTIT, 1990

Une remarque pour conclure cette discussion sur la définition des normes. On trouve beaucoup d'espèces de régularités dans une société comme la nôtre. Il y a des lois, par exemple, ce sont des régularités que l'Etat soutient. Il y a des conventions, par exemple les conventions qui règlent la langue, le commerce et la vie quotidienne. Intuitivement, il devrait être possible qu'une loi ou une convention soit en même temps une norme. Ce que je voudrais souligner, en concluant, c'est que la définition que j'ai proposée préserve cette possibilité. En effet, il est possible, selon la définition ci-dessus, qu'une loi soit une norme. Une loi sera aussi une norme si elle remplit nos trois conditions, ce qui est évidemment possible. En effet, il est possible qu'une même régularité soit soutenue par l'Etat et que, simultanément, le fait de l'approbation rende la régularité plus probable qu'autrement. Il en va de même pour la convention. David Lewis (3) a spécifié son concept de convention à l'aide de trois conditions, que ma définition reprend. Une régularité est une convention si et seulement si

1. Presque tout le monde se conforme,

2. Presque tout le monde anticipe que presque tout le monde s'y conforme,

3. Cette anticipation est pertinente pour la conformité – presque tout le monde préfère se conformer si tous les autres se conforment, parce que la conformité universelle résout un problème de coordination.

Mais, il est évidemment possible qu'une régularité soit soutenue pour une part par cette pression conventionnelle et que le fait de l'approbation rende la conformité plus probable qu'autrement. Cette dernière observation distingue mon approche de celle de Jon Elster. Selon Elster, une régularité n'est une norme que si les gens se conforment à cette régularité en raison de motivations particulières. Il représente le comportement normatif comme une espèce particulière de comportement. Le

résultat est que Elster ne peut pas préserver l'intuition que j'ai exposée (4). Sous cet angle, mon approche est supérieure.

Qu'est-ce que la théorie du choix rationnel ?

Nous pouvons traiter brièvement cette seconde question. Pour notre propos, la définition contenue dans une remarque de John Harsanyi suffit : « People's behavior can be largely explained in terms of two dominant interests – economic gain and social acceptance ». Ce que Harsanyi soutient, c'est que les gens sont en général, dans une certaine mesure, égoïstes : ils poursuivent leurs intérêts personnels. Il distingue deux sortes d'intérêt personnel : l'avantage économique et l'acceptation sociale. Il n'en dit pas plus sur le sens de cette distinction, mais nous pouvons la comprendre de la manière suivante : l'avantage économique est celui que l'on gagne par des actions, des actions de soi-même ou des actions des autres. L'acceptation sociale est une autre espèce d'avantage. C'est l'avantage offert, non par l'action, mais par l'attitude, en l'occurrence, par les attitudes favorables des autres : être apprécié, admiré, aimé, ou n'importe quel autre type d'acceptation. On peut acheter l'avantage économique, du moins en général, logiquement il n'y a pas de problème si on paie les autres pour qu'ils vendent ce qu'on voudrait posséder. Mais on ne peut pas acheter l'acceptation sociale. On ne peut que mériter les attitudes favorables des autres, ou, au moins, faire semblant de les mériter. Il n'est pas possible que quelqu'un offre sincèrement d'avoir une attitude favorable si on le paie pour obtenir cette attitude. La logique empêche le « marché attitudinal ».

Il y a trois raisons pour lesquelles quelqu'un pourrait rejeter cette présentation de la théorie du choix rationnel. La première est que cette présentation ne mentionne pas la rationalité elle-même

(3) LEWIS, 1969

(4) ELSTER, 1991

Harsanyi dit que l'intérêt explique en grande partie le comportement humain, mais il ne dit pas que les hommes poursuivent leurs intérêts de manière rationnelle. Cette remarque est juste. Il faut supposer la rationalité pour comprendre ce que dit Harsanyi, en l'occurrence, la rationalité instrumentale. Les gens poursuivent leurs intérêts par un type de comportement qui maximise leur satisfaction anticipée – disons, par un type de comportement qui est rationnel vis-à-vis de leurs intérêts, et ce dans le sens instrumental de la rationalité tel que la théorie bayésienne l'explique.

Une deuxième raison pour laquelle on pourrait rejeter la caractérisation de Harsanyi est qu'elle suppose l'hégémonie de l'égoïsme. Beaucoup d'économistes, et d'autres théoriciens, disent qu'ils supposent que les gens sont rationnels, au sens de la théorie de la décision, sans pour autant supposer que les gens n'ont que des désirs égoïstes. En réponse à cet argument, je dis d'abord que Harsanyi suppose que les désirs égoïstes sont importants, il ne suppose pas qu'il n'y a pas d'autres désirs. Et je dis de plus que, malgré leur affirmation et leur méthodologie, les économistes supposent, lorsqu'ils développent leurs explications, que dans l'action les gens sont égoïstes. C'est pour cette raison qu'ils supposent en général que les préférences que les hommes manifestent en action sont des préférences qu'on doit satisfaire comme utilitaristes. Amartya Sen a éclairé cette supposition dans quelques articles (5).

La troisième raison pour laquelle on pourrait rejeter la caractérisation de Harsanyi n'est pas plus convaincante que la deuxième. Harsanyi, semble-t-il, identifie le désir d'acceptation comme un désir distinct et indépendant du désir de l'avantage économique. Beaucoup de gens, au moins beaucoup d'économistes, soulignent que le bien lié au fait d'être accepté serait en fait celui d'avoir par ce moyen une meilleure chance de jouir de l'avantage économique. Il y a plusieurs choses que je voudrais dire en réponse. Mais je les dirai plus loin.

Normes et choix rationnels

Venons-en à notre propos central : la théorie du choix rationnel peut-elle expliquer les normes ou, au moins, expliquer les normes qui nous concernent, et notamment, les normes générales qui servent à résoudre les problèmes collectifs ? Mais que veut dire « expliquer les normes » ? Ce que j'entends par cette expression, c'est expliquer non pas pourquoi des normes ont émergé dans une société – c'est là une question pour l'histoire – mais pourquoi elles continuent d'être suivies, pourquoi elles se maintiennent.

La théorie du choix rationnel va pouvoir expliquer le maintien en vigueur de certaines normes dans une société si elle peut identifier chez les membres de la société des intérêts personnels qui les motivent à se comporter d'une façon telle qu'ils assurent la persistance des normes pertinentes. Cette explication est une explication par la main invisible. Dans la société, les gens se comportent d'une certaine façon, chacun ayant ses raisons personnelles d'agir comme il le fait, et le résultat de l'agrégation de leurs comportements – quelque chose que personne peut-être n'a anticipé – est que certaines normes règnent dans la société.

J'ai noté que David Lewis a développé une définition des conventions qui est très bien connue des théoriciens du choix rationnel. Il a lui aussi tenté d'expliquer les conventions à partir des intérêts personnels et même d'expliquer des conventions qui sont des normes – des conventions qui existent, en partie, à cause de l'approbation du comportement conforme. Son explication est aussi une explication par la main invisible.

Lewis montre qu'il y a de bonnes raisons personnelles pour que tout le monde se comporte dans les termes requis par une convention – de bonnes raisons qui ne supposent pas elles-mêmes que la convention existe – et en même temps de bonnes raisons pour que tout le monde approuve la conformité des autres et désapprouve la

(5) Par exemple SEN, 1979

déviations. Il suppose que si tout le monde approuve ou désapprouve en ces termes, alors cela rendra la conformité plus probable. Nous pouvons dire alors non seulement que les conventions existent, mais aussi qu'elles constituent des normes : presque tout le monde se comporte selon les régularités en question, presque tout le monde approuve la conformité et désapprouve la déviation, et ce fait rend la conformité plus probable que le cas contraire.

Pourquoi, dit Lewis, tout le monde doit-il se comporter selon les modalités qu'une convention requiert ? Il suppose que le comportement en question, pour n'importe quelles raisons, est le comportement anticipé chez les autres, et que les situations pertinentes font surgir un problème de coordination. Si on ne se comporte pas en ces termes, alors on ne va pas se coordonner avec les autres. Pourquoi, dit Lewis, tout le monde doit-il approuver la conformité et désapprouver la déviation, dans cette situation ? Approuver la conformité de quelqu'un, c'est être dans la disposition d'en faire l'éloge, et désapprouver c'est être dans la disposition de le censurer – plus exactement, être dans la disposition de le faire, au moins, dans une situation où l'éloge ou la censure est accessible. Lewis dit que tout le monde devrait approuver la conformité, et devrait désapprouver la déviation, parce que la conformité universelle est non seulement un équilibre – cela veut dire que personne ne peut dévier avec bénéfice – mais aussi un équilibre de coordination : personne ne peut dévier sans nuire à l'autre.

Les normes conventionnelles, telles que Lewis les explique, sont des normes exceptionnelles. En général, les normes ne sont pas des conventions. Elles ne satisfont pas la définition de Lewis. Étant donné une convention, tout le monde a une raison personnelle de s'y conformer s'il croit que tout autre va s'y conformer. Mais cela ne vaut pas pour la plupart des normes, même celles qui résolvent des problèmes collec-

tifs. Rappelez-vous les normes citées plus haut : dire la vérité, tenir ses promesses, ne pas voler, remplir ses devoirs civiques, contribuer aux projets collectifs. Avec ces normes, tout le monde a une raison personnelle de dévier s'il croit que tout un chacun va se conformer, tout le monde peut bénéficier de la conformité des autres sans se conformer soi-même.

Le problème pour la théorie du choix rationnel est alors d'expliquer, et d'expliquer par rapport aux intérêts personnels, pourquoi les gens se conforment aux normes non conventionnelles : aux normes telles que nous les avons illustrées. Si nous cherchons une explication par la main invisible, une explication du type de celle que Lewis nous offre pour les normes conventionnelles, nous devons expliquer pourquoi les gens se comportent de manière pertinente et pourquoi ils approuvent la conformité et désapprouvent la déviation, nous pouvons supposer, comme Lewis le suppose, que cette approbation et cette désapprobation, si elles existent, vont rendre la conformité plus probable que dans le cas contraire.

L'idée clé de la théorie du choix rationnel, telle qu'elle a été développée dans les vingt dernières années, est que nous pouvons expliquer les normes non conventionnelles par la disposition rationnelle à faire *tit-for-tat* : à rendre à chacun la pareille (6). Pour que nous puissions comprendre cette approche, il faut expliquer pourquoi cette disposition est rationnelle et comment le comportement normatif est produit par cette disposition. Voici en bref l'argument que les théoriciens développent (j'en commenterai les différents niveaux).

1 *Tit-for-tat* est une stratégie rationnelle dans une série illimitée de dilemmes d'échange entre deux partenaires,

2 Pour les mêmes raisons *tit-for-tat* est une stratégie rationnelle dans une série illimitée de dilemmes d'échange entre plusieurs partenaires,

3 Les situations pour lesquelles une norme est pertinente constituent une série

(6) On pourrait aussi traduire *tit-for-tat* par « donnant-donnant », comme dans la version française du livre de R. AXELROD, 1984.

illimitée de cette espèce les problèmes de *free riding* (faire cavalier seul) sont des dilemmes d'échange ,

4 Dans ce cas nous pouvons expliquer le comportement coopératif – le comportement que la norme requiert – comme une réponse rationnelle ,

5 Le résultat de la coopération est un équilibre de coordination dans lequel personne ne peut changer son comportement sans nuire aux autres, et, étant donné le comportement coopératif, tout le monde doit approuver la conformité et désapprouver la déviation ,

6 Il faut que cette approbation et cette désapprobation rendent la conformité plus probable et que la régularité constitue une norme

La rationalité de la disposition à rendre la pareille surgit du rôle qui est la sienne dans le dilemme du prisonnier ou, plus strictement, dans une série illimitée de dilemmes d'échange Qu'est-ce que le dilemme du prisonnier ? L'histoire invoquée en général, et qui explique le dilemme, est la suivante Le chef de la police vient d'arrêter deux hommes pour un crime Le policier sait que les hommes ont commis le crime mais il sait aussi qu'il ne va pas pouvoir l'établir au cours du procès Il construit un dilemme pour chaque prisonnier dans le but de le convaincre d'avouer le crime

Ce qu'il dit à chacun c'est la chose suivante « Tu peux avouer ou nier que toi et ton copain avez fait le crime Mais voici une bonne raison d'avouer ton copain va avouer ou nier le crime et j'ai arrangé les choses de telle manière qu'il est assuré que l'aveu est la meilleure option pour toi, que l'autre avoue le crime ou le nie Supposons que l'autre nie En ce cas, tu seras libre, et l'autre restera cinq ans en prison Mais si tu nies, tu passeras un an en prison, avec ton copain, car je vous accuserai, tous les deux, d'avoir résisté à l'arrestation Supposons inversement que l'autre avoue En ce cas, si tu avoues aussi, tu ne resteras que trois ans en prison, avec ton copain, mais si tu nies, tu resteras cinq ans en prison »

Dans le dilemme du prisonnier, les récompenses et les amendes sont structurées

d'une façon qui rend l'aveu rationnel Mais l'aveu est une stratégie non coopérative et le résultat coopératif – le déni par les deux – est plus avantageux les prisonniers passeront un an en prison et pas trois Donc ce qui est individuellement rationnel pour les deux prisonniers n'est pas avantageux pour eux Ils serviraient mieux leurs intérêts s'ils pouvaient coopérer l'un avec l'autre en niant le crime Mais la rationalité empêche la coopération Il est rationnel de trahir l'autre, et non rationnel de coopérer avec lui

La théorie du choix rationnel accepte que la non-coopération est rationnelle dans le dilemme du prisonnier, alors que la coopération entre les prisonniers est plus avantageuse pour eux que la non-coopération Ce paradoxe est au centre de la théorie On a montré qu'il y a beaucoup de situations sociales ressemblant au dilemme du prisonnier et le modèle est souvent cité dans la littérature Le dilemme est en général appelé « dilemme de l'échange » On le rencontre quand on trouve deux partenaires dont chacun a des raisons décisives – des raisons qui sont indépendantes du choix de l'autre – de trahir l'autre, alors que la coopération de la part des deux est supérieure à la non-coopération Ou, selon une définition moins stricte, on rencontre le dilemme d'échange quand on trouve deux partenaires dont aucun ne peut tirer avantage de la coopération, quel que soit le choix de l'autre la non-coopération est la seule stratégie équilibrée, alors que la coopération de la part des deux est plus avantageuse pour chacun que la déviation des deux

Supposons que la théorie du choix rationnel soit juste et que la non-coopération soit rationnelle dans le dilemme de l'échange Ce résultat fait surgir beaucoup de questions Ces questions tournent autour du fait surprenant que les gens en général peuvent coopérer, et que souvent en coopérant ils réussissent à obtenir des bénéfices d'échange En explorant ces questions la théorie a dirigé notre attention sur ce qui doit se passer dans les situations où les dilemmes d'échange ne sont pas isolés En particulier, elle a dirigé l'attention sur la situation où on trouve une série illimitée

de dilemmes d'échange. Qu'est-ce que les gens doivent faire ? Est-ce qu'ils doivent se trahir chaque fois ? Ou est-ce qu'ils doivent adopter une autre stratégie ?

La stratégie de défection permanente – la stratégie de trahison – a le caractère d'un équilibre si les deux la suivent, alors aucun ne peut rejeter la stratégie avec avantage. Mais ce qui est surprenant avec une série illimitée de dilemmes d'échange, c'est qu'il y a beaucoup de stratégies équilibrées. Par exemple, la stratégie de *tit-for-tat* – faire la pareille – ou la stratégie de double *tit-for-tat* ou la stratégie de *tit-for-double-tat*. Faire *Tit-for-tat* cela veut dire coopérer au premier tour et puis faire ensuite ce que l'autre aura fait au tour précédent. Aucun ne peut rejeter une telle stratégie avec bénéfice si elle est établie entre les deux.

Mais s'il y a beaucoup de stratégies équilibrées dans une série illimitée de dilemmes d'échange, ces stratégies ne sont pas de même statut. La stratégie de *Tit-for-tat*, par exemple, est supérieure à la stratégie de défection permanente. *Tit-for-tat* est un exemple de stratégie de coopération conditionnelle et avec une telle stratégie les deux parties s'en sortent mieux qu'avec la défection permanente. La stratégie est supérieure au sens de Pareto : sous cette stratégie, au moins l'un va mieux et l'autre ne va pas pis.

Le résultat de ces réflexions est que chaque partie doit rationnellement chercher une stratégie conditionnelle de coopération, si du moins elle s'intéresse aux résultats des dilemmes à venir. Mais quelle stratégie ? La chose la plus importante c'est qu'elle adopte la même stratégie que l'autre, sinon il y a le danger qu'elles ne puissent pas coordonner leurs actions. Nous sommes ainsi confrontés à la question que Lewis posait en discutant les conventions.

Dans un livre remarquable, Robert Axelrod (7), s'inspirant de l'œuvre de Michael Taylor (8) entre autres, a montré

qu'il y a beaucoup de raisons pour que chacune des parties adopte *tit-for-tat*, c'est-à-dire choisisse de rendre la pareille (9). On ne peut pas prouver la supériorité de la stratégie *more geometrico*, mais Axelrod a montré que cette stratégie a des avantages que les autres n'ont pas. Il a organisé un concours de stratégies différentes par ordinateur, dans lequel *tit-for-tat* a réussi et il a analysé les raisons de ce succès. La disposition à rendre la pareille est une disposition rationnelle dans une série illimitée de dilemmes d'échange.

La rationalité de la stratégie est facile à comprendre. C'est une stratégie telle que les partenaires, dans une série de dilemmes d'échange, peuvent anticiper mutuellement les décisions et ainsi se coordonner. C'est une stratégie avec des bénéfices manifestes pour les deux parties. Et c'est une stratégie croyable. Si un partenaire dit qu'il va coopérer dans le premier tour et ensuite faire ce que l'autre aura fait dans le tour précédent, l'autre peut le croire, c'est une menace rationnelle, parce qu'un partenaire rationnel ne doit pas continuer à coopérer si l'autre ne coopère pas.

La rationalité de la stratégie est soutenue, d'après Axelrod, par le fait que dans le monde naturel beaucoup d'organismes sont portés à la suivre, et par le fait qu'on la trouve souvent dans le monde social. Il mentionne un exemple frappant de la Première Guerre mondiale. Les soldats dans les tranchées ont régulièrement mis en œuvre cette stratégie. Ils tiraient au-dessus des têtes des soldats du camp adverse sauf quand ceux-ci leur tiraient dessus. La coordination était formidable, et, en général, c'était une coordination par la stratégie de *tit-for-tat*. C'était un grave problème pour les officiers que de convaincre les soldats de tirer sur l'ennemi. Ils essayaient de briser la coordination en faisant faire des mouvements de troupes, mais cela n'avait aucun résultat. Ils n'ont pu réussir qu'en 1917, lorsqu'ils ont introduit les raids de nuit, dont le seul but était de briser la coor-

(7) AXELROD, 1984

(8) TAYLOR, 1987

(9) Cf HARDIN, 1982, SUGDEN, 1986

dination entre les soldats des parties adverses

Alors, disons que la disposition à faire *tit-for-tat* est une disposition rationnelle dans une série illimitée de dilemmes d'échange. Mais quelle est sa pertinence pour expliquer le comportement normatif ? Ce que beaucoup de théoriciens du choix rationnel ont soutenu ces dernières années c'est qu'on peut expliquer ce comportement par la disposition à faire *tit-for-tat*. Ils considèrent que les situations pour lesquelles les normes dont nous parlons sont pertinentes sont des situations comme le dilemme d'échange. Lorsque ces situations présentent une série illimitée de dilemmes d'échange, il est rationnel pour les partenaires de faire *tit-for-tat*, il est rationnel pour eux de coopérer d'abord – de dire la vérité, de tenir ses promesses, etc – et ensuite de faire ce que les autres auront fait au tour précédent.

Pourquoi les situations pour lesquelles les normes sont pertinentes sont-elles comme des dilemmes d'échange ? Pour comprendre ce que ces théoriciens veulent dire, nous devons généraliser la notion de dilemme d'échange. Deux partenaires se trouvent dans un dilemme de cette espèce quand la seule stratégie équilibrée est de trahir l'autre, mais que la coopération de la part des deux est supérieure à leur non-coopération. Nous généralisons cette définition en disant que plusieurs partenaires se trouvent dans un dilemme d'échange si la seule stratégie équilibrée est la stratégie non coopérative et si la coopération universelle est supérieure, pour tous, à la non-coopération universelle. Etant donné cette définition, nous pouvons comprendre pourquoi les théoriciens disent que les situations pour lesquelles les normes sont pertinentes sont des dilemmes d'échange. Considérons une situation où tout le monde a le choix entre dire la vérité et dire ce qui lui convient le mieux. Chacun a des raisons personnelles de dire ce qui lui convient le mieux, quel que soit le choix des autres – c'est la meilleure option pour chacun si les autres disent la vérité en gé-

néral, si les autres disent en général ce qui leur convient le mieux ou si les autres ne suivent pas en général la même stratégie. Mais bien qu'il ne soit pas rationnel de dire la vérité, tout le monde préfère que ce choix soit universel. La coopération universelle est supérieure à la déviation universelle. Cette explication soulève beaucoup de questions, mais je suis prêt à admettre dans le contexte présent que les situations pour lesquelles les normes sont pertinentes sont comme des dilemmes d'échange (10).

Pourquoi la stratégie de *tit-for-tat* est-elle une bonne stratégie pour les gens rationnels, face à une série illimitée de dilemmes d'échange de cette espèce ? Nous devons généraliser la stratégie pour plusieurs personnes – chacune coopère au premier tour et ne coopère dans les tours suivants que si toutes les autres ont coopéré au tour précédent. Cette stratégie est une stratégie équilibrée et elle promet un meilleur résultat que la stratégie de la déviation permanente. Elle est équilibrée parce que si tout le monde met en œuvre cette stratégie, personne ne peut tirer avantage d'un changement de comportement. Cette stratégie promet un meilleur résultat que la stratégie de déviation parce que si tout le monde la suit, alors tout le monde bénéficiera de la coopération.

Nous avons vu que les théoriciens du choix rationnel défendent les points suivants

1 La non-coopération est rationnelle dans un dilemme d'échange à deux partenaires ,

2 La disposition à faire *tit-for-tat* – une stratégie conditionnelle de coopération – est rationnelle dans une série illimitée de tels dilemmes. Elle est aussi équilibrée que la disposition à dévier, elle promet un meilleur résultat et, parmi les stratégies équilibrées de cette espèce, elle est la plus frappante ,

3 Les situations pour lesquelles les normes sont pertinentes sont des dilemmes d'échange à plusieurs parties, mais il s'agit d'une série illimitée de dilemmes ,

4 La stratégie généralisée de *tit-for-tat* est rationnelle dans une série de cette espèce, pour les mêmes raisons que dans le cas des dilemmes à deux partenaires ,

5 Alors nous devons anticiper que les gens, en tant qu'acteurs rationnels, coopéreront dans les situations pour lesquelles les normes sont pertinentes , cela veut dire que nous devons anticiper que les gens vont se conformer aux normes

Nous en avons presque fini avec la présentation de la théorie des normes que les théoriciens du choix rationnel soutiennent. Ils ont montré que les gens vont mettre en œuvre le comportement requis par une norme universelle non conventionnelle. Mais pour qu'une norme existe, il faut que les gens approuvent le comportement conforme et désapprouvent le comportement non conforme. Les théoriciens du choix rationnel peuvent-ils expliquer cette approbation et cette désapprobation ? Si on suit l'approche de David Lewis, on dira qu'ils le peuvent. La conformité à la stratégie de *tit-for-tat* est, comme son nom l'indique, un équilibre : personne ne peut tirer bénéfice d'un changement de comportement. Mais cette conformité est aussi un équilibre de coordination au sens de Lewis (11) : personne ne peut changer son comportement sans nuire aux autres , si quelqu'un le fait, les autres doivent le punir par la déviation et la déviation a un coût

Parce que la conformité à la stratégie de *tit-for-tat* est un équilibre de coordination, tout le monde doit être disposé, dans une situation convenable, à censurer ceux qui dévient et à faire l'éloge de ceux qui se conforment (12). Mais cela veut dire que tout le monde approuve la conformité et désapprouve la déviation. J'ai dit au commencement de cette discussion que j'étais prêt à admettre que l'approbation et la désapprobation ont des conséquences : elles rendent plus probable que les gens se conforment à la régularité en question. Nous voyons maintenant comment les théoriciens du choix rationnel espèrent expliquer les normes. La disposition à faire

tit-for-tat explique la conformité à la régularité. Et, étant donné la conformité générale, on peut expliquer pourquoi nous trouvons une norme : l'approbation et la désapprobation sont intelligibles et elles doivent soutenir la conformité générale.

LES NORMES ET LA MAIN INTANGIBLE

Les théoriciens du choix rationnel suggèrent donc de considérer les normes comme les produits d'une main invisible. Les acteurs rationnels suivent la stratégie généralisée de *tit-for-tat* dans les situations pour lesquelles les normes dont nous parlons – les normes générales, positives et non conventionnelles – sont pertinentes. Cela veut dire qu'ils se trouvent dans la disposition d'approuver la conformité à la norme et de désapprouver la déviation. Les régularités auxquelles la conformité est produite par la stratégie de *tit-for-tat* constituent des normes : presque tout le monde approuve la conformité et désapprouve la déviation et, bien sûr, cette approbation et cette désapprobation rendent la conformité plus probable. Les normes sont établies dans la société sans que personne ne l'ait prévu ni ne l'ait désiré.

Les limites de l'explication orthodoxe

Il y a vingt ans, personne ne savait comment expliquer la conformité aux normes non-conventionnelles dans la théorie du choix rationnel. Cette proposition récente représente une tentative importante pour combler cette lacune. Mais la tentative a échoué, les théoriciens cités n'ayant pas pris en considération un problème important. Ils ont montré que le résultat dans lequel chacun suit *tit-for-tat* est un équilibre et, plus spécifiquement, un équilibre de coordination : personne ne peut changer son comportement avec bénéfice et personne ne peut changer son comportement sans nuire aux autres. Ils ont aussi montré que

(11) HARDIN, 1982

(12) SUGDEN, 1986

le résultat est meilleur que l'équilibre produit par la stratégie de déviation permanente et, bien sûr, qu'il est aussi bon que tous les autres équilibres le meilleur résultat est la coopération universelle et le *tit-for-tat* universel produit la coopération universelle Mais cela ne suffit pas pour notre propos Il faut montrer non seulement que le résultat de *tit-for-tat* est un équilibre, mais aussi que la stratégie elle-même représente un choix rationnel Il est important que le résultat du *tit-for-tat* universel soit le plus avantageux possible, mais cela ne montre pas encore que *tit-for-tat* est un choix rationnel

Voici une raison pour laquelle il n'est pas un choix rationnel Nous savons que, dans une série illimitée de dilemmes d'échange, beaucoup de stratégies équilibrantes peuvent produire une coopération universelle Un exemple en est la stratégie de double *tit-for-tat*, un autre la stratégie de triple *tit-for-tat* Il faut qu'on puisse montrer que *tit-for-tat* est la meilleure parmi ces stratégies équilibrantes Les théoriciens du choix rationnel ont soutenu que *tit-for-tat* est plus frappant, comme stratégie, que d'autres stratégies conditionnelles, et que les partenaires rationnels doivent se coordonner sur le *tit-for-tat* Acceptons que *tit-for-tat* soit plus frappant – et de ce fait plus attrayant pour les acteurs rationnels – que les autres stratégies conditionnelles de coopération On n'a pas montré pour autant qu'il constitue un choix rationnel

Il y a une autre raison pour laquelle une stratégie comme *tit-for-tat* n'est peut-être pas un choix rationnel Il est possible qu'une stratégie produise un équilibre si tout le monde la suit, et cela même si la stratégie n'est pas accessible comme choix rationnel Peut-être l'équilibre se présente-t-il comme une contrée attrayante à laquelle personne ne peut accéder

Une stratégie rationnelle doit remplir trois conditions

1 Le résultat de la conformité est un équilibre ,

2 Le résultat est rationnellement accessible ,

3 Le résultat est le meilleur équilibre accessible

Il faut donc qu'on puisse montrer que *tit-for-tat* est accessible comme choix rationnel pour les gens qui se trouvent dans les situations pour lesquelles les normes dont nous parlons sont pertinentes Il importe, de ce point de vue, que la stratégie soit la plus frappante parmi les stratégies coopératives conditionnelles Mais, tout en étant la stratégie la plus frappante de toutes, *tit-for-tat* peut être inaccessible comme choix rationnel

Je prétends que *tit-for-tat* est inaccessible dans les situations pertinentes La stratégie fait en effet problème dans le cas suivant Le partenaire qui est dans la disposition de ne pas se conformer peut-il croire les autres quand ils disent qu'ils vont suivre la stratégie de *tit-for-tat*, qu'ils vont coopérer au premier tour, puis coopérer ensuite si et seulement si tous les autres ont coopéré au tour précédent ? Je soutiens que celui qui est dans la disposition de dévier ne croira pas les autres quand ils adopteront *tit-for-tat* Le *tit-for-tat* généralisé n'est pas une stratégie croyable

Pourquoi ? Parce que la menace de répondre à une seule défection par la défection générale n'est pas croyable Le déserteur potentiel que je viens d'imaginer peut raisonner de la façon suivante Si les autres coopèrent et si je fais défection, les autres m'en voudront, mais ils n'auront pas recours à la menace impliquée par *tit-for-tat* En effet, appliquer la menace, cela veut dire retourner à la situation où personne ne coopère, et cette situation – la déviation universelle – est beaucoup moins avantageuse pour les autres que la situation où tout le monde, moi excepté, coopère Le problème est réel

Pourquoi les théoriciens du choix rationnel n'ont-ils pas remarqué ce problème ? Deux raisons à cela La première est que les économistes tendent en général à supposer que dès lors qu'une stratégie s'avère la seule stratégie équilibrante ou la meilleure stratégie équilibrante possible, l'acteur rationnel va la choisir La seconde est que la stratégie de *tit-for-tat* n'est accessible – en particulier croyable – que dans un dilemme d'échange limité à deux partenaires Dans cette situation chacun a en effet avantage,

si l'autre dévie, à dévier lui-même. Il est donc avantageux pour chacun d'appliquer la menace impliquée par *tit-for-tat* et chacun peut croire l'autre quand il annonce qu'il va suivre cette stratégie. Ce fait explique peut-être pourquoi les théoriciens du choix rationnel n'ont pas remarqué que *tit-for-tat* n'est ni croyable, ni accessible, dans les dilemmes d'échange impliquant un nombre de partenaires supérieur à deux.

L'approche par la main intangible

Cependant le fait que l'approche par la main invisible ne réussisse pas à expliquer les normes ne permet pas de montrer que la théorie du choix rationnel est en défaut. Il existe une autre approche de la théorie, que je vais présenter. On ne la trouve pas dans la littérature de la théorie du choix rationnel, mais du point de vue de la théorie, elle représente une approche intéressante et peut-être importante.

Nous avons défini les normes comme des régularités que l'approbation et la désapprobation rendent plus probables. L'approche par la main invisible tente d'expliquer d'abord le comportement conforme, puis l'approbation et la désapprobation, elle suppose que si l'approbation et la désapprobation existent, celles-ci vont rendre la conformité plus probable. Mais notre définition des normes nous dirige vers une autre possibilité. Elle suggère que nous devrions peut-être d'abord postuler la présence d'*attitudes* pertinentes, puis expliquer le comportement conforme en fonction de ces attitudes.

Nous avons vu, d'après la définition de John Harsanyi, que la théorie du choix rationnel postule que les acteurs rationnels poursuivent soit l'avantage économique, soit l'acceptation sociale. Bénéficiaire de l'acceptation sociale c'est être apprécié, admiré ou aimé, bref susciter l'approbation. Mais comment peut-on expliquer la conformité aux normes pertinentes en termes de recherche de l'acceptation so-

ciale ? L'idée clé est qu'en offrant l'acceptation sociale en plus de l'avantage économique, les partenaires peuvent faire surgir des bénéfices que personne n'a anticipés. Cette explication est une explication par la main intangible.

La tradition orthodoxe a ignoré l'explication de la conformité aux normes en fonction de fins d'acceptation sociale. Plus exactement, elle a accepté un argument qui a l'air d'avoir détruit tout effort explicatif de ce genre. James Buchanan, un économiste qui a reçu le Prix Nobel, a présenté cet argument sous la forme d'un dilemme, le dilemme de la contrainte – *the « enforcement dilemma »* (13). Son idée est que nous aurions tort de croire que l'application d'une norme en fonction de l'approbation et de la désapprobation pourrait résoudre un dilemme d'échange, si les gens sont confrontés à un dilemme d'échange, ils sont confrontés à un dilemme identique quand ils cherchent une solution normative.

Si les gens sont confrontés à un dilemme d'échange – ne serait-ce qu'à un problème de *free-riding* – cela implique qu'ils ne sont pas très motivés. Peut-être se soucient-ils du bien-être de la société en soi, mais ils sont aussi motivés par des raisons personnelles. Il n'est donc pas certain qu'ils se conforment aux régularités normatives dont nous traitons ici. Mais si, de plus, ils ne sont pas parfaitement moraux, s'ils sont motivés par des raisons mélangées, il n'est pas certain qu'ils appliquent les régularités normatives pertinentes du fait de l'approbation et de la désapprobation. L'application de même que la conformité coûtent, dit Buchanan : « Enforcement has two components. First, violations must be discovered and violators identified. Second, punishment must be imposed on violators. Both components involve costs » (14). Donc, si nous n'anticipons pas la conformité générale à une régularité, nous ne devons pas non plus anticiper l'application de la régularité : nous ne devons pas anticiper la conformité à la méta-régularité d'appliquer la régularité originale.

(13) BUCHANAN, 1975

(14) Id., p. 132-133

Devons-nous accepter cette objection ? Bien que la tradition orthodoxe l'ait fait sien depuis longtemps, il est évident que l'argument n'est pas valide. Buchanan suppose que l'application par l'approbation et la désapprobation ne peut que prendre la forme d'une action intentionnelle. Ce sont seulement les actions intentionnelles qui suscitent les pénalisations que Buchanan a citées. Ce qu'il envisage c'est que les gens surveillent intentionnellement leurs actions mutuelles, faisant l'éloge du comportement conforme aux normes et censurant le comportement non conforme. Mais approuver et désapprouver n'impliquent pas nécessairement de faire l'éloge et de censurer. On approuve le comportement d'un autre quand on est disposé à en faire l'éloge, et cela dans une situation où l'éloge ne pose pas problème (par exemple, il n'entraîne ni l'embarras, ni aucune difficulté de ce genre). Inversement, on désapprouve le comportement quand on est disposé à le censurer, et cela dans une situation où la censure est possible – elle n'entraîne pas de pénalisations sérieuses. Mais, si les situations ne sont pas appropriées, on peut approuver ou désapprouver sans rien faire intentionnellement. Et, bien sûr, on peut identifier les gens qui se conforment et les gens qui dévient sans les surveiller intentionnellement.

L'application d'une norme en fonction d'attitudes – son application en termes d'approbation ou de désapprobation quand ces attitudes n'impliquent pas d'action intentionnelle – ne permet pas les pénalisations que Buchanan cite. C'est pourquoi son argument n'est pas valide. En principe l'idée que les gens font surgir la conformité aux normes par une attitude d'approbation ou de désapprobation ne fait pas difficulté. Peut-être leurs motifs moraux n'assurent-ils pas la conformité aux normes. De même est-il possible qu'ils approuvent la conformité et désapprouvent la déviation et que, par ce moyen, ils s'offrent, les uns aux autres, un motif personnel de se conformer. Il est possible qu'ils appliquent une norme, qu'ils assurent que

presque tout le monde s'y conforme, sans rien faire et sans rien dire.

Arrivés à ce point, nous pouvons comprendre pourquoi il est utile de faire une distinction entre la main invisible et la main intangible. Il s'agit d'une main invisible quand les acteurs font des actions intentionnelles mais qu'ils ne produisent pas intentionnellement leur résultat pertinent, les actions peuvent produire ce résultat sans que personne l'ait anticipé ou sans que personne l'ait désiré. Il s'agit d'une main intangible quand les acteurs ne font aucune action intentionnelle, au moins au premier tour, qui entraînerait la sanction des autres. Dans ce cas, les acteurs produisent un bénéfice par leurs attitudes non-intentionnelles et non par leurs actions non intentionnelles (15).

Mais la tradition orthodoxe peut répondre à cette suggestion que nous avons tort de supposer que les acteurs rationnels recherchent une acceptation sociale qui se limiterait à la présence d'attitudes inexprimées. Elle peut objecter que l'approbation n'est attrayante pour des acteurs rationnels, et la désapprobation pénible, que si elles sont exprimées. Si les gens appliquent une norme en fonction de l'approbation et de la désapprobation, il faut qu'ils expriment l'une ou l'autre de ces attitudes quand ils voient les autres se conformer ou dévier. Mais s'ils doivent se comporter ainsi, alors Buchanan a raison : il y a un dilemme de l'application.

Trois possibilités doivent pourtant être distinguées dans ce cas.

1 L'approbation des autres est attrayante pour un acteur rationnel si et seulement si les autres l'expriment,

2 L'approbation des autres peut être attrayante pour un acteur rationnel, même s'ils ne l'expriment pas, parce qu'elle implique qu'ils sont dans la disposition de bien agir avec lui,

3 L'approbation des autres est attrayante pour un acteur rationnel, même s'ils ne l'expriment pas, parce que, *ceteris paribus*, l'approbation est un bien intrinsèque (et la désapprobation un mal intrinsèque).

L'approche par la main intangible suppose que la première thèse est fautive et que l'une des deux autres est vraie. Cette supposition est convaincante. Nous supposons, dans la théorie du choix rationnel, que les acteurs poursuivent au moins l'avantage économique. Il est évident que l'on accroît la chance que les autres se conduisent bien avec soi lorsque l'on gagne leur approbation. C'est pourquoi un acteur rationnel doit rechercher l'approbation, même l'approbation que personne n'exprime au moment de l'action. Adam Smith le dit : « We are pleased to think that we have rendered ourselves the natural objects of approbation, though no approbation should ever actually be bestowed upon us, and we are mortified to reflect that we have justly merited the blame of those we live with, though that sentiment should never actually be exerted against us » (16).

Cette observation permet de répondre à l'objection de la tradition orthodoxe. Mais comment choisir entre la thèse 2 et la thèse 3 ? Je voudrais aborder cette question de la manière suivante : y a-t-il de bonnes raisons pour croire que les gens aiment l'approbation et détestent la désapprobation, comme un bien et un mal intrinsèques ? Voici trois considérations pertinentes par rapport à cette question.

La première est que ce postulat est lié à une longue tradition qui considère que les hommes ont le désir fondamental d'être acceptés, estimés, appréciés, aimés ou de jouir de n'importe quelle espèce d'acceptation. Adam Smith lui-même appartient à cette tradition. Il a cru que ce qui explique en partie le désir des biens économiques c'est que la richesse implique la distinction et l'approbation.

La seconde considération est qu'il y a beaucoup de preuves que les hommes recherchent l'acceptation sociale, sans que pour autant cette acceptation ait un lien avec le gain économique. Tout le monde déteste être observé, même par des étrangers, quand on fait des choses embarrassantes. Et beaucoup de gens font des

choses qui ne rapportent qu'un honneur posthume.

Finalement, il y a des raisons liées à la théorie de la sélection naturelle qui font que les gens doivent regarder l'acceptation sociale comme un bien intrinsèque. Supposons que l'avantage économique ait une importance intrinsèque pour les hommes, à cause de son importance biologique. Supposons aussi que l'acceptation soit importante comme moyen d'acquiescer cet avantage économique. Étant donné ces suppositions – et elles sont assez convaincantes – nous pouvons reconnaître que peut-être les hommes sont sélectionnés, par la nature, aux fins de rechercher non seulement l'avantage économique mais aussi l'acceptation sociale. Il est probable que la nature ait anticipé la rationalité économique et cela, avant que les individus puissent débattre du lien entre l'acceptation et l'avantage. Il est probable que la nature ait engendré chez les hommes un désir fondamental d'être approuvés par les autres.

Si nous regardons l'acceptation sociale comme un bien intrinsèque pour les gens – un bien aussi fondamental, disons, que l'avantage économique – nous pouvons admettre qu'il est possible que les hommes recherchent chaque bien sous deux angles : comme un bien intrinsèque, et comme un moyen d'acquiescer d'autres biens, c'est-à-dire comme un bien instrumental. L'acceptation sociale est un moyen, nous l'avons vu, d'accéder à l'avantage économique. Et l'avantage économique, comme Smith l'a souligné, est aussi un moyen de renforcer l'acceptation sociale.

L'approche par la main intangible est pertinente si les hommes recherchent l'acceptation sociale comme un bien intrinsèque. Mais l'approche garde aussi une réelle pertinence dans le cas où l'acceptation sociale n'est qu'un bien instrumental. Je vais supposer, dans la discussion qui suit, que l'acceptation est un bien intrinsèque. Même si vous n'acceptez pas ce postulat, vous pouvez néanmoins accepter, dans certaines limites, ce que je vais soutenir.

L'explication des normes par la main intangible

Reprenons l'argument principal. Nous avons vu que l'approche par la main intangible, telle que je viens de la décrire, peut être aussi justifiée, en principe, que l'approche par la main invisible. L'objection posée par la tradition orthodoxe de la théorie du choix rationnel n'est pas probante. Mais si l'approche peut en principe être justifiée, est-elle utile en ses conséquences ?

Je pense que oui (17). Nous pouvons distinguer deux hypothèses selon lesquelles cette approche peut expliquer, au moins en partie, la conformité aux normes. La première est faible, la deuxième est plus radicale.

La première suppose que les gens sont naturellement disposés à approuver un comportement parce qu'ils en voient les effets positifs sur le bien-être du groupe auquel ils appartiennent, et, inversement, qu'ils sont disposés à désapprouver les comportements qui ont des effets négatifs sur ce groupe. Cette hypothèse n'appartient pas à la théorie du choix rationnel. Mais on peut l'étayer sur le postulat que les hommes font l'expérience de la sympathie au sens de Hume et de Smith, ou sur des postulats formulés par des sociologues tels que Parsons ou même Bourdieu : ceux de l'intériorisation des valeurs ou de l'habitus. Supposons alors que les conditions suivantes soient remplies dans une société

1. Presque tout le monde croit que la conformité à une régularité R, dans une situation S, fait progresser le bien-être général et que la déviation le fait régresser ;

2. Presque tout le monde croit que la première supposition est vraie ;

3. Presque tout le monde croit qu'il y a une grande chance d'être identifié selon que l'on se conforme à la régularité ou que l'on en dévie ;

4. La conformité ne coûte pas beaucoup

Si ces conditions sont remplies, les ac-

teurs rationnels vont se trouver dans une situation où la conformité à R va susciter l'approbation et la déviation la désapprobation. C'est pourquoi, il est bien possible que la main intangible produise, ou aide à produire, la conformité à la régularité. Et si la main intangible est pertinente de cette manière, la régularité est une norme : presque tout le monde se conforme à la régularité, presque tout le monde approuve la conformité des autres et désapprouve la déviation, et ce fait est pertinent – il rend la conformité plus probable qu'autrement.

Les théoriciens du choix rationnel seront-ils heureux du rôle que leur théorie joue dans le cadre de cette hypothèse faible ? Je pense que ce rôle est assez important et que l'hypothèse que je viens de décrire est probante pour certaines normes. C'est une hypothèse implicite chez les philosophes classiques et modernes. Mais je doute que les théoriciens du choix rationnel se réjouissent du rôle ainsi assigné à leur théorie. En général, quand ils défendent leur théorie et en font l'éloge, ces théoriciens lui donnent un rôle plus important.

Venons-en à une hypothèse plus radicale, une hypothèse pour laquelle la théorie du choix rationnel peut jouer un rôle plus important encore. C'est l'hypothèse que l'acceptation sociale, qui contribue à engendrer la conformité à certaines régularités, ne peut être produite elle-même, en général, sans un mécanisme en rapport avec la théorie du choix rationnel. On suppose, dans la perspective radicale, que la théorie du choix rationnel peut expliquer pourquoi les hommes approuvent la conformité et désapprouvent la déviation. Alors la théorie aurait une double pertinence. Elle pourrait expliquer, au moins en partie, pourquoi les hommes approuvent la conformité à certaines normes et désapprouvent la déviation, et pourquoi ils se conforment à ces normes.

Nous savons qu'un homme approuve ou désapprouve quelque chose quand il est disposé à en faire l'éloge ou à le censurer, étant donné une situation dans laquelle

(17) PETTIT, 1990

cette réponse est accessible – elle n’implique pas de pénalités. Supposons toujours que l’approbation puisse être fondée sur des raisons personnelles et non seulement sur des raisons morales, par exemple, des raisons qui portent sur le bien-être de la société. Alors, pourquoi des acteurs rationnels devraient-ils approuver la conformité à certaines régularités et désapprouver la déviation ?

L’acteur rationnel va approuver la conformité d’un autre, et désapprouver sa déviation, quand la conformité – même un seul acte de conformité – lui fait du bien et quand la déviation – même un seul acte de déviation – lui fait du mal. Etant donné un tel acte, l’acteur rationnel sera disposé à en faire l’éloge ou à le censurer dans une situation où ces réponses ne provoquent pas de pénalités. Par ce moyen, l’acteur rationnel peut renforcer ou inhiber le comportement.

Supposons, alors, que les conditions suivantes sont remplies

1* Presque tout le monde croit que la conformité, même un seul acte de conformité, à une régularité R, dans une situation S, fait progresser son bien-être personnel et que la déviation le fait régresser ,

2 Presque tout le monde croit que la première supposition est vraie ,

3 Presque tout le monde croit qu’il y a une grande chance d’être identifié, que l’on se conforme à la régularité R dans une situation S, ou que l’on dévie ,

4 La conformité ne coûte pas beaucoup

Avec ces hypothèses, presque tout le monde reconnaîtra que la conformité à la régularité provoquera l’approbation des autres et la déviation leur désapprobation. Donc, avec ces hypothèses, la main intangible produira l’approbation et la désapprobation pertinentes. Il est possible aussi que, par ce moyen, la main intangible produise, ou aide à produire, la conformité elle-même. Elle produira la conformité, on le suppose, quand la conformité n’implique pas de grandes pénalités – quand elle ne coûte pas beaucoup. Les acteurs rationnels accorderont l’acceptation sociale à un comportement conforme et, en l’absence de fortes pénalités, la recherche de cette acceptation sociale les conduira à se conformer.

Nous avons supposé jusqu’ici que l’approbation et la désapprobation peuvent être fondées sur des raisons morales, que l’attitude qui est pertinente face aux régularités normatives est de l’ordre de l’approbation morale. La théorie du choix rationnel peut-elle expliquer, au moins en partie, pourquoi les gens approuvent et désapprouvent de cette manière ?

Qu’est-ce qu’approuver ou désapprouver pour des raisons morales ? C’est être dans la disposition, étant donné une situation pertinente, de faire l’éloge ou de censurer, et de faire l’éloge ou de censurer par rapport à des considérations morales par rapport, par exemple, à des considérations relevant du bien-être général. Bref, c’est être dans la disposition de moraliser, de faire de la morale, au moins dans une situation où la moralisation n’implique pas de pénalité. La question que je viens de poser est alors la suivante : la théorie du choix rationnel peut-elle expliquer la disposition à moraliser, en particulier la disposition à moraliser au sujet de la conformité à certaines normes ?

La réponse doit être affirmative. Si les dernières conditions sont remplies, nous pouvons voir que les acteurs rationnels vont approuver non seulement la conformité, mais aussi la moralisation qui est produite dans le but de faire progresser la conformité. Et si ces conditions sont remplies, nous pouvons voir également que les acteurs rationnels vont désapprouver, non seulement la déviation, mais la moralisation qui est produite dans le but de faire régresser la déviation. Avec ces conditions pertinentes, la théorie du choix rationnel, en particulier l’explication par la main intangible, s’annonce excellente.

Nous avons vu ainsi que la théorie du choix rationnel peut, pour traiter les normes, se servir de l’explication par la main intangible sous certaines conditions. Sous les trois premières conditions, elle peut offrir une explication modeste par la main intangible. Elle suppose que l’acceptation sociale est produite par un mécanisme étranger à la théorie, mais elle l’invoque pour expliquer la conformité à certaines normes. Sous les autres conditions elle peut offrir une explication plus

radicale une explication non seulement de la conformité à certaines normes, mais aussi de l'approbation et de la désapprobation – de l'acceptation sociale – qui aident à produire cette conformité

Pour conclure, voici une liste d'exemples de normes que la théorie du choix rationnel pourrait expliquer, semble-t-il, soit de façon modeste, soit de façon radicale

- la norme de la vaine pâture ,
- la norme de faire la queue et de prendre la file ,
- la norme d'être consciencieux lorsqu'on fait partie d'un comité – par exemple, d'un comité de nomination, d'un comité d'organisation ou d'un jury ,
- la norme de ne pas fumer dans un lieu public

Pour chacun de ces exemples nous devons poser deux questions

- la théorie du choix rationnel peut-elle expliquer pourquoi presque tout le monde approuve la conformité et désapprouve la déviation ? La condition 1* est-elle remplie ?

- la théorie du choix rationnel peut-elle expliquer pourquoi ces attitudes contribuent à produire la conformité ? Les conditions 1-2-3-4 ou 1*-2-3-4 sont-elles remplies ?

Si nous pouvons répondre à ces deux questions pour des normes existantes, alors nous pourrions identifier, d'une façon plus générale, les conditions qui contribuent à l'établissement des normes et celles qui les font disparaître. Cette observation montre comment la théorie pourrait être pertinente pour la philosophie politique et, plus généralement, pour le discours éthique. Si nous recommandons une constitution pour la société – si nous recommandons une constitution pour n'importe quelle institution – nous supposons que des normes pertinentes peuvent être établies. Mais comment défendre une telle supposition ? Comment soutenir que certaines normes sont « faisables », comme disent les économistes ? Si la théorie du choix rationnel peut répondre en général à des questions du genre des précédentes, elle nous offrira

un moyen par lequel nous pourrions peut-être identifier des normes « faisables » ou praticables

L'EXPLICATION PAR LES CHOIX RATIONNELS N'A-T-ELLE AUCUNE VALIDITÉ ?

Après avoir critiqué l'explication orthodoxe des normes – la stratégie de *tit-for-tat* qu'elle invoque ne fonctionne pas avec plusieurs partenaires – , nous avons défendu la théorie du choix rationnel en suggérant qu'elle peut peut-être expliquer les normes autrement que par la main invisible, à savoir par la main intangible. L'idée est que la poursuite de l'acceptation sociale – la recherche de l'approbation et le refus de la désapprobation – produit, ou contribue à produire, la conformité à certaines normes. Les gens ne se sanctionnent pas intentionnellement mais non intentionnellement. Chacun se comporte en tenant compte des attitudes qu'il voit surgir chez les autres en réaction à ce qu'il fait et chacun recherche, *ceteris paribus*, des attitudes favorables. Ces attitudes ne sont pas produites intentionnellement – cela n'est pas possible – mais elles peuvent jouer un rôle de sanction. Elles peuvent peut-être expliquer, et expliquer selon les termes de la théorie du choix rationnel, la conformité à certaines normes.

En expliquant l'existence des normes par la main intangible, nous avons supposé que la théorie du choix rationnel est valide en général. Il est vrai que le comportement des gens se prête à une explication par les intérêts personnels, qu'il s'agisse d'intérêts économiques ou d'intérêts sociaux. Nous allons maintenant lever cette supposition et nous demander si la théorie, en particulier en ce qu'elle est une théorie des normes, est valide ou pas.

La position que je vais défendre sur cette question est une position œcuménique ou pluraliste. Ce que je crois, c'est que la théorie du choix rationnel a une certaine validité mais que sa validité est limitée, et que la théorie permet à d'autres

théories – à des théories apparemment antagonistes – d’avoir une validité parallèle (18)

Intérêts personnels et raisons impersonnelles

Il n’y a pas de problème à supposer qu’on peut expliquer en grande partie le comportement des gens vis-à-vis du marché par la théorie du choix rationnel. Mais le marché n’est pas typique. Il ne représente pas la situation normale. Tout le monde croit qu’en situation de marché les gens poursuivent leurs intérêts personnels – au moins dans une société comme la nôtre – mais tout le monde croit, par ailleurs, qu’ils ne se comportent pas en général comme s’ils étaient en situation de marché. Le marché ne représente ni la situation des gens chez eux, ni leur situation en politique, ni leur situation sociale en général. Il ne représente pas, en particulier, la situation des gens face aux normes. Quand des acteurs ordinaires sont confrontés aux normes, la plupart ne considèrent pas leurs intérêts personnels. Ils reconnaissent ce que la norme requiert d’eux et en général s’y conforment sans réfléchir. Et s’ils ne se conforment pas, ça n’est pas toujours pour avoir réfléchi aux coûts de la conformité, la faiblesse de la volonté ne constitue pas toujours une réussite de la rationalité personnelle !

Donc tout le monde sait que le marché n’est pas typique et que les gens se comportent souvent selon d’autres considérations que les raisons personnelles. Mais cette thèse n’est pas seulement une thèse de la sociologie du peuple. Le verdict de la psychologie sociale, de l’anthropologie ethnographique, de la criminologie – le verdict presque universel des sciences sociales – est que les gens ne se comportent pas en général à la manière de l’*homo oeconomicus*. Ces sciences sociales rejoignent le sens commun, lorsqu’elles disent que les hommes sont motivés par des raisons diverses : des raisons d’amour, de dévotion, de fidélité, de devoir. Ces raisons portent non seulement sur leurs intérêts personnels mais aussi sur les intérêts des autres, sur les intérêts des groupes aux-

quels ils appartiennent et, bien sûr, sur l’intérêt commun.

Cette hypothèse est difficile à nier, au moins dans une perspective qui traite les êtres humains comme des sujets intentionnels, des sujets qui se comportent d’une manière qui se prête à une explication en termes de croyances et de désirs. Bien sûr, beaucoup de gens ont l’air de ne considérer que des raisons impersonnelles, alors qu’ils se réfèrent essentiellement à leurs propres intérêts. Par exemple, bien qu’ils ne mentionnent que des raisons impersonnelles au fondement de leurs actions, beaucoup de politiciens ne cherchent qu’à assurer leur avenir électoral. Mais admettre cette possibilité n’est pas nier l’hypothèse présentée, cette possibilité va de pair avec l’hypothèse selon laquelle les gens introduisent des raisons impersonnelles aussi bien que personnelles dans leurs délibérations, et que leurs considérations ne sont pas nécessairement inefficaces : elles peuvent affecter ce qu’ils font.

Il est difficile de nier cette hypothèse. Pourtant, la théorie du choix rationnel n’en a tenu aucun compte, bien qu’elle ne l’ait pas niée explicitement. On peut expliquer cette inattention par une ambiguïté que nous avons identifiée plus haut entre deux hypothèses : celle que les acteurs rationnels font preuve de rationalité instrumentale, au sens de la théorie Bayésienne, et celle, plus spécifique, que les acteurs rationnels font preuve non seulement de rationalité en ce sens, mais encore d’une rationalité égoïste. Selon cette dernière hypothèse, les acteurs rationnels n’exercent la rationalité instrumentale qu’en poursuivant leurs intérêts personnels. Les théoriciens du choix rationnel, me semble-t-il, n’invoquent pratiquement, pour expliquer les actions, que les intérêts personnels, alors qu’ils se déclarent prêts, en principe, à admettre d’autres motivations et même à les invoquer. Leur comportement explicatif n’est pas en accord avec leur méthodologie explicative.

Pourquoi cette ambiguïté permet-elle aux théoriciens du choix rationnel de ne pas tenir compte de l’hypothèse citée ? Les théoriciens peuvent répondre à la difficulté que l’hypothèse fait surgir par rapport à

leur méthodologie, en l'ignorant Cherchant à expliquer le comportement humain, ils traitent la théorie du choix rationnel comme une théorie égoïste Mais ils peuvent aussi prendre l'hypothèse en considération Et ils présentent la théorie comme une théorie générale de la rationalité comme une théorie instrumentale, bien sûr, et non comme une théorie égoïste

Pourquoi dire que la théorie du choix rationnel n'invoque que les intérêts personnels pour expliquer le comportement humain ? Je vois trois raisons pertinentes La première est que les théoriciens du choix rationnel ne citent eux-mêmes que les raisons personnelles Bien sûr, la théorie de la décision, la théorie de la rationalité en soi, ne suppose pas que les acteurs rationnels sont des égoïstes Mais la théorie appliquée soutient cette supposition pratiquement

La deuxième raison est que ce trait de la théorie appliquée n'est pas un accident La théorie ne peut pas offrir d'explications intéressantes hors de doctrines telles que l'égoïsme Si elle ne suppose pas l'égoïsme, si elle admet n'importe quelle espèce de motivation, alors elle peut expliquer tout ce que les hommes font, elle peut l'expliquer sans difficulté, et sans hasard, par rapport à tous les désirs – peut être celui-ci, peut-être celui-là – que le comportement satisfait ou tend à satisfaire Mais une telle entreprise ne présente aucun intérêt intellectuel Il faut que la théorie du choix rationnel impose une restriction, telle celle que l'égoïsme représente, aux motivations qu'elle se permet d'invoquer

La troisième raison pour laquelle je dis que la théorie du choix rationnel n'invoque que des intérêts personnels se rapporte à une observation de Amartya Sen Sen a remarqué deux choses pertinentes D'abord que les préférences qu'un utilitariste voudrait satisfaire ne peuvent être que des préférences personnelles si nous traitons d'une société dans laquelle il y a des égoïstes et des altruistes, l'utilitariste ne va pas recommander la satisfaction égale des désirs égoïstes et altruistes, cela ne serait pas juste, car cela punirait les altruistes

Sen a remarqué, en deuxième lieu, que les économistes supposent en général que les préférences qu'ils invoquent pour expliquer le comportement humain sont des préférences qu'on doit satisfaire en tant qu'utilitariste Qu'en conclure ? La conclusion la plus raisonnable est que les économistes supposent, malgré ce qu'ils disent en tant que méthodologues, que les préférences qu'ils invoquent pour expliquer le comportement humain sont des préférences égoïstes

Retournons à l'hypothèse que nous avons discutée, celle selon laquelle les gens se comportent en général en fonction de considérations diverses qui peuvent affecter ce qu'ils font et non en fonction de leurs seuls intérêts personnels Qu'est-ce que la théorie du choix rationnel doit répondre à cette hypothèse ? Il y a trois possibilités Les deux premières impliquent soit de rejeter, soit de réviser l'hypothèse Peu de théoriciens ont identifié la troisième possibilité, alors qu'elle est cohérente tant avec l'hypothèse qu'avec la théorie du choix rationnel

La première possibilité est de rejeter l'hypothèse en disant que, malgré les apparences, les hommes se conduisent en considérant leurs intérêts personnels Ils délibèrent en eux-mêmes sur les récompenses et les pénalités personnelles des différentes options auxquelles ils se trouvent confrontés et en général ils se décident à faire ce qui fera progresser leurs intérêts Les hommes se vouent explicitement à la recherche de leur satisfaction personnelle Cette approche n'est pas populaire et il est difficile de lui trouver un adhérent, même un seul Les théoriciens du choix rationnel n'aiment pas nier la phénoménologie de la délibération telle qu'elle relève du sens commun

La deuxième possibilité est de dire que les acteurs rationnels se consacrent implicitement, et non explicitement, au développement de leurs intérêts personnels Cette approche reconnaît que les hommes introduisent des considérations diverses dans leurs délibérations et qu'ils donnent l'impression de se décider sur la base de ces considérations Mais il ne faut pas se

fier aux apparences. Alors que les gens introduisent des considérations diverses, les seules considérations qui sont vraiment puissantes sont les raisons personnelles. Gary Becker soutient ce point de vue, me semble-t-il. « The economic approach does not assume that decision units are necessarily conscious of their own efforts to maximize or otherwise describe in an informative way reasons for the systematic patterns in their behaviour. Thus it is consistent with the emphasis on the subconscious in modern psychology » (19)

Cette deuxième possibilité est plus attrayante que la première, parce qu'elle nous permet d'admettre que les hommes ont l'apparence – bien sûr, l'apparence trompeuse – d'être motivés par des raisons impersonnelles, elle nous permet d'admettre que la phénoménologie de la délibération est valide comme description du niveau conscient. Mais cette possibilité n'est ni très attrayante ni très convaincante. Nous voudrions dire non seulement que les hommes donnent l'impression de tenir compte des raisons impersonnelles, mais que cette impression est souvent vraie, en réalité les hommes tiennent compte à la fois de raisons impersonnelles et de raisons personnelles. Nous n'avons aucune raison de croire que ce que les acteurs ont l'air de faire, ce qu'ils font au niveau conscient, n'est pas en vérité ce qui se passe.

Le scénario des intérêts virtuels

Venons-en à la troisième réponse possible. Cette réponse est implicite, comme nous le verrons, dans l'approche de David Lewis quand il tente d'expliquer la conformité aux conventions. Mais personne ne l'a remarquée parmi les méthodologues et les philosophes de la théorie du choix rationnel. La réponse admet que les intérêts personnels ne sont pas réellement présents – ni explicitement, ni implicitement – dans les délibérations qui précèdent beaucoup d'actions, y compris des actions pour les-

quelles la théorie du choix rationnel est pertinente. Elle dit cependant que les intérêts personnels sont virtuellement présents dans de telles délibérations et que cela suffit à assurer la pertinence de la théorie du choix rationnel.

Qu'est-ce que ça veut dire que les intérêts personnels peuvent être le fait d'une présence virtuelle, par opposition à une présence réelle soit explicite, soit implicite ? L'idée clé est la suivante. Les acteurs peuvent se conduire au vu de seules raisons impersonnelles, c'est-à-dire en fonction de buts qui n'impliquent pas, au moins directement, leurs propres intérêts. Mais si leur comportement dessert leurs intérêts personnels, par rapport à un certain modèle, cela déclenche un signal d'alarme et ils vont commencer à introduire des raisons personnelles parmi les considérations sur lesquelles se fonder. En outre, ces raisons vont produire un effet, du moins chez la plupart des acteurs, et le comportement va mieux servir leurs propres intérêts. Les intérêts personnels sont virtuellement présents en ce sens qu'ils sont toujours un élément accessible dans la délibération si le comportement les fait régresser. Ils sont comme des contraintes de réserve que le comportement doit satisfaire. Des contraintes qui vont s'insinuer dans la délibération si le comportement ne permet pas de les satisfaire.

La possibilité que j'ai à l'esprit est indéterminée, car je n'ai rien dit sur le modèle pertinent de la satisfaction des intérêts personnels. Supposons qu'un acteur rationnel se comporte en ne tenant compte que de considérations impersonnelles. Supposons qu'il ne fasse pas progresser ses intérêts personnels. A quel point le signal d'alarme va-t-il se déclencher ? A quel moment la lampe témoin s'allumera-t-elle ? Avant de répondre à cette question, je voudrais suggérer que la possibilité en question ici est plausible, même dans une forme indéterminée. Nous savons tous que les intérêts personnels sont importants. On a réalisé depuis plus de deux mille ans que le grand

(19) BECKER 1976

problème posé par un ordre social, un ordre moral, est la contrainte imposée aux intérêts personnels leur élimination, la réduction de leur influence ou leur réorientation dans une direction donnée. Mais nous savons aussi que les gens ne sont pas des égoïstes ils se conduisent en fonction de raisons impersonnelles autant que personnelles, ils recherchent le bien-être des autres, celui des groupes auxquels ils appartiennent autant que le leur propre. Le scénario que nous avons décrit donne leur place à ces deux propositions. Les raisons invoquées peuvent être personnelles ou impersonnelles – à cet égard elles sont d'une importance égale – mais les raisons personnelles bénéficient toujours d'une présence virtuelle et dans cette mesure elles sont primaires. D'ailleurs le scénario qui postule des intérêts virtuels fait leur place aux deux propositions sans être obligé de recourir à des hypothèses relevant d'un niveau psychologique inconscient.

Quand va se déclencher le signal d'alarme dont nous avons parlé ? Il est bien possible qu'à cet égard il y ait des différences entre les individus et entre les cultures. Mais je voudrais montrer comment en principe on pourrait répondre à la question. Voici une hypothèse possible. Tout le monde appartient à un groupe primaire de référence, à un groupe qui détermine ses idéaux et ses espérances, en lui donnant un modèle pour la vie. Les liens qui déterminent le groupe de référence d'un individu sont multiples. Les membres d'un groupe participent à une culture commune et sont de même statut économique, social et légal. Le signal d'alarme va s'allumer, selon mon hypothèse, lorsque l'individu va réaliser qu'il ne peut pas atteindre, à cause de son comportement, les mêmes bénéfices que ses camarades de groupe.

Qu'arrive-t-il ? Peut-être les autres dans le groupe ne suivent-ils pas les normes auxquelles il conforme son comportement. Les autres ne payent pas les impôts qu'il doit payer par exemple. Ou peut-être a-t-il rejoint un nouveau groupe, parce qu'il a gagné beaucoup d'argent et que son comportement habituel n'a plus de

sens par rapport à ceux de ses camarades de groupe. Il y a beaucoup de possibilités.

Nous avons introduit une possibilité – le scénario des intérêts virtuels – selon laquelle les acteurs rationnels peuvent se comporter en tenant compte des raisons impersonnelles alors que les raisons personnelles font l'objet d'un statut spécial qui est le suivant. Les acteurs ne veulent pas voir leurs intérêts personnels frustrés. Mais pour que la théorie du choix rationnel soit valide – la théorie au sens d'Harsanyi – il faut pouvoir expliquer ce que les gens font par rapport à leurs intérêts personnels. Nous n'avons pas encore expliqué ce qui est possible de ce point de vue. Nous allons donc considérer cette question.

Le scénario des intérêts virtuels peut se comparer au scénario suivant. On a une balle qui roule en ligne droite, la ligne droite symbolise le comportement qui fait progresser les intérêts personnels. La balle roule en ligne droite à cause de lois générales – les lois de Newton par exemple – et non pas à cause de facteurs spéciaux, ces lois symbolisent les raisons impersonnelles de même que les facteurs spéciaux représentent les raisons personnelles. Mais si la balle est perturbée, et si elle cesse de rouler en ligne droite, alors des facteurs spéciaux sont activés, qui vont rectifier son trajet. Ces facteurs vont mettre fin aux perturbations et assurer que la balle retourne à la ligne droite. Nous pouvons imaginer ces facteurs spéciaux comme des bornes qui se trouvent disposées sur les deux côtés de la ligne.

Cette analogie éclaire la question de savoir comment, dans le scénario des intérêts virtuels, les intérêts personnels peuvent expliquer le comportement. Ceux-ci correspondent aux bornes. Mais les bornes ne peuvent pas expliquer pourquoi la balle roule en ligne droite, quand il n'y a pas de perturbations. Elles ne causent rien, elles ne sont pas des causes actuelles mais des causes de réserve. Nous devons invoquer les lois de Newton, et non pas des facteurs spéciaux, pour expliquer pourquoi la balle roule en ligne droite, ces lois nous orientent vers les facteurs pertinents. Plus précisément encore, on ne doit pas invoquer ces

facteurs spéciaux pour expliquer la course de la balle, ils ne sont pas pertinents

La leçon est claire. Les raisons impersonnelles sont les seules raisons pertinentes du comportement qu'elles produisent. Mais ce comportement est contraint de satisfaire d'une manière ou d'une autre les raisons personnelles, celles-ci constituent alors des causes de réserve qui assurent que le comportement ne va pas léser les intérêts personnels. Nous sommes pourtant confrontés à un problème. La théorie du choix rationnel peut-elle avoir une pertinence explicative dans le scénario des intérêts virtuels ? Ne nous sommes-nous pas trompés en introduisant ce scénario ?

Non, je ne crois pas. Les bornes ne peuvent expliquer ni l'émergence ni la continuation du comportement de la balle. Par analogie, les intérêts personnels ne peuvent expliquer ni l'émergence ni la continuation du comportement des gens dans la situation envisagée. Mais les bornes peuvent rendre compte d'une autre caractéristique du comportement de la balle. De même qu'elles peuvent expliquer pourquoi ce comportement est résilient – pourquoi la balle roulera en ligne droite, quelle que soit la perturbation (20) – de même les intérêts personnels, alors qu'ils n'ont qu'une présence virtuelle, peuvent expliquer pourquoi le comportement des gens sera assez résilient – pourquoi il conservera sa forme, même si les motivations actuelles commencent à orienter l'acteur dans une autre direction.

Tout le monde reconnaît l'émergence et la continuation d'un événement ou d'un processus – d'un comportement, par exemple – comme un *explanandum* approprié pour les sciences sociales. Peu de méthodologues ont remarqué que la résilience est aussi un *explanandum* approprié (21). Pourtant il est évident que les sciences sociales ont souvent affaire à cette espèce d'*explanandum*. Elles essaient d'expliquer non seulement pourquoi certains événements se sont passés, mais aussi pourquoi

ces événements étaient inévitables, inéluctables. De ce point de vue, nous voyons ici une différence entre les sciences sociales et l'histoire, du moins l'histoire telle qu'on la retrouve chez les historiens traditionnels. L'histoire traditionnelle recherche les causes particulières des événements, tandis que les sciences sociales recherchent les causes générales – les causes qui, quand elles existent, rendent un certain résultat inévitable, quoi qu'il arrive par ailleurs.

La même chose va souvent expliquer l'occurrence d'un événement et son caractère inévitable. Mais l'analogie de la balle montre qu'il est possible qu'il y ait des différences entre les deux aspects. Les lois de Newton expliquent le fait que la balle roule en ligne droite. Les bornes expliquent la nécessité ou la résilience qui s'attache à ce fait. Dans la situation sociale qui est pertinente pour nos propos, les raisons impersonnelles expliquent le fait que les hommes se conduisent d'une certaine manière – celle qui fait progresser leurs intérêts personnels. Quant aux raisons personnelles, elles expliquent la résilience qui s'attache à ce fait.

Je parle toujours de résilience et d'inévitabilité sans autre qualification. À vrai dire, il serait préférable de parler de résilience ou d'inévitabilité relatives. Il n'est pas absolument inévitable que la balle roule en ligne droite quelle que soit la protection contre la perturbation. De même, il n'est pas absolument inévitable que les gens, quelle que soit leur situation, se comportent d'une façon qui permette de satisfaire leurs intérêts personnels. Mais la résilience relative peut constituer un *explanandum* aussi intéressant que la résilience absolue.

On peut se demander, à propos de cette résilience, soit relative, soit absolue, si nous pouvons l'observer avant d'avoir une explication pour le phénomène qu'elle constitue. Nous pouvons comprendre ce qu'est la résilience ou l'inévitabilité. Et nous pouvons juger, étant donné certains signes, qu'un événement, un processus ou

(20) La résilience est plus que la récurrence ou la régularité : elle est de l'ordre de la résistance naturelle en fonction d'une fécondité ou d'une productivité (note de L. Q.)

(21) PETTIT 1993

n'importe quoi bénéficient de la résilience. On doit cependant admettre que nous n'en sommes pas sûrs avant d'avoir trouvé une explication appropriée. Mais cela ne fait aucun problème. La situation est parallèle à celle qu'on trouve avec une théorie générale et les lois qu'elle permet d'expliquer. Il se peut que nous ne soyons pas sûrs qu'une régularité est une loi ou pas, avant d'avoir identifié une explication théorique pour cette loi. Mais nous n'hésitons pas à dire dans une telle situation que la théorie explique la loi.

Le point de vue que je voudrais défendre est, enfin, accessible. Je suis d'accord avec les théoriciens qui critiquent la théorie du choix rationnel quand ils disent que les gens ne se conforment pas au modèle de l'*homo oeconomicus*. Je suis d'accord sur l'argument que les gens ne sont ni des égoïstes explicites, ni des égoïstes implicites. Mais je crois quand même que la théorie du choix rationnel peut être pertinente à propos du comportement humain, et qu'elle peut être pertinente même à propos du comportement qui est le produit de raisons impersonnelles. Il se peut qu'elle permette d'expliquer non pas pourquoi les gens se comportent de fait d'une telle manière, mais pourquoi, de cette manière, leur comportement est assez résilient.

A partir de ce point de vue, qu'est-ce que les théoriciens du choix rationnel doivent faire ? Ils doivent essayer d'identifier des types de comportement qui, étant donné un contexte social, ont l'air d'être assez résilients et rechercher les résultats d'ensemble que ces types de comportement font surgir. L'assurance que ce programme puisse aboutir n'est pas donnée *a priori*. Même si le modèle des intérêts virtuels est valide, il se peut qu'on ne trouve pas au niveau d'abstraction requis des types résilients de comportement, et même si on identifie des types appropriés, il se peut qu'ils ne fassent surgir aucun résultat intéressant. Je ne défends ici que la possi-

bilité que la théorie du choix rationnel puisse constituer, en dehors du marché, un programme significatif.

Retournons brièvement à la question des normes. J'ai présenté plus haut une explication de la conformité aux normes dans les termes de la théorie du choix rationnel. Est-ce que nous pouvons expliciter cette tentative comme l'essai d'expliquer la résilience des normes pertinentes ? Je prétends que oui. David Lewis reconnaît que les conventions qu'il essaie d'expliquer par la théorie du choix rationnel ne sont pas des produits du raisonnement : les gens se conforment par habitude plutôt qu'en fonction d'une réflexion sur les bénéfices de la conformité qu'il invoque lui-même dans son explication. Alors quelle est la pertinence de la théorie du choix rationnel par rapport à la conformité ? Lewis suggère une réponse – une réponse qui concorde avec le scénario des intérêts virtuels – quand il fait la remarque suivante sur le caractère d'habitude de la conformité : « If that habit ever ceased to serve the agent's desires according to his beliefs, it would at once be overridden and corrected by conscious reasoning » (22).

Nous pouvons comprendre de la même manière la pertinence de notre explication des normes. Celle-ci montre que la conformité servira les intérêts personnels des gens qui suivent les normes en question : elle suscitera l'acceptation sociale. La conformité elle-même peut être l'aboutissement du raisonnement impersonnel, de l'habitude, ou de n'importe quoi. La chose importante est que la conformité serve bien les intérêts personnels de sorte que nous puissions dire qu'elle va survivre même si les gens commencent à raisonner d'une autre façon ou s'ils changent leurs habitudes. C'est pourquoi la théorie du choix rationnel, en particulier l'explication par la main intangible, peut rendre compte, sinon du fait de la conformité, du moins de sa résilience.

RÉFÉRENCES

- AXELROD, R , 1984, *The Evolution of Cooperation*, New York, Basic Books (trad fr *Donnant-donnant*, Paris, Ed Odile Jacob, 1992)
- AXELROD, R , 1986, « An Evolutionary Approach to Norms », *American Political Science Review*, 80 (4), p 1095-1111
- BECKER, G , 1976, *The Economic Approach to Human Behavior*, Chicago, University of Chicago Press
- BRENNAN, G et PETTIT, P , 1993, « Hands Invisible and Intangible », *Synthèse*, 94, p 191-225
- BUCHANAN, J , 1975, *The Limits of Liberty*, Chicago, University of Chicago Press
- ELSTER, J , 1991, « Rationality and Social Norms », *Archives européennes de sociologie*, XXXII, p 109-129
- JACKSON, F et PETTIT, P , 1993, « Pour l'œcuménisme explicatif », in P Ladrrière, P Pharo et L Quéré (EDS), *La théorie de l'action*, Paris, CNRS Editions, p 23-52
- HARDIN, R , 1982, *Collective Action*, Baltimore, Johns Hopkins University Press
- HARSANYI, J , 1969, « Rational Choice Models of Behavior versus Functionalist and Conformist Theories », *World Politics*, 22, p 513-538
- LEWIS, D , 1969, *Convention*, Cambridge, Mass , Harvard University Press
- LEWIS, D , 1983, « Languages and Language », in *Id* , *Philosophical Papers*, I, Oxford, Oxford University Press (trad fr partielle *supra*)
- PETTIT, P , 1986, « Free Riding and Foul Dealing », *Journal of Philosophy*, 83, p 361-379
- PETTIT, P , 1989, « Foul Dealing and an Assurance Problem », *Australian Journal of Philosophy*, 67, p 341-344
- PETTIT, P , 1990, « Virtus Normativa Rational Choice Perspectives », *Ethics*, 100, p 725-755
- PETTIT, P , 1993, *The Common Mind an Essay on Psychology, Society and Politics*, New York, Oxford University Press
- SEN, A , 1979, « Rational Fools », in F Hahn et M Hollis (eds), *Philosophy and Economic Theory*, Oxford, Oxford University Press
- SMITH, A , 1982, *The Theory of Moral Sentiments*, Indianapolis, Liberty Classics
- SUGDEN, R 1986, *The Economics of Rights Cooperation and Welfare*, Oxford, Basil Blackwell
- TAYLOR, M , 1987, *The Possibility of Cooperation*, Cambridge, Cambridge University Press