Commentaires sur J. Elster, « L'indétermination de la théorie du choix rationnel »

M. Cozic (IHPST, GREGHEC & DEC)

La contribution de Jon Elster porte sur les différentes formes d'indétermination de la théorie du choix rationnel. Plus précisément, elle traite principalement de *l'indétermination dans la formation des croyances rationnelles*, croyances qui peuvent porter sur l'environnement « passif » - dans ce cas il parle d'incertitude « brute » - ou sur l'environnement « actif », c'est-à-dire les autres agents – et dans ce cas il parle d'incertitude « stratégique ». Il y a trois types d'indétermination ou d'incertitude qui sont considérées :

1°/ L'incertitude concernant le comportement d'agents avec qui on doit interagir

2°/ L'incertitude concernant la forme de la distribution de certaines variables qui ont une importance particulière pour des décisions (par exemple, les variables climatiques dans les questions actuelles de politique environnementale)

3°/ L'incertitude concernant la recherche optimale d'information (quand faut-il arrêter de chercher de l'information ?)

Il ne m'est pas possible, dans les commentaires qui vont suivre, de rendre justice à la richesse des thèmes abordés par J. Elster. Je vais me concentrer sur quatre points auxquels mes propres recherches m'ont rendu particulièrement sensible.

(C1) La première question que je voudrais poser est essentiellement une question de clarification : c'est celle de savoir en quel sens l'indétermination constitue un *échec* pour la théorie du choix rationnel. Une comparaison peut aider à préciser la question. La logique, à laquelle on compare parfois la théorie du choix rationnel – R. Jeffrey a intitulé sa célèbre monographie consacrée à la théorie de la décision *The Logic of Decision*, ne prescrit pas ce que l'on doit croire de manière déterminée. Elle interdit tout au plus d'avoir certaines croyances – des croyances incohérentes. Si l'on considère la théorie du choix rationnel comme un ensemble de normes, alors il n'y a rien de surprenant ni de très problématique à ce que cet ensemble ne prescrive pas systématiquement le choix de telle action ou l'adoption de telles croyances. On répondra peut-être que l'indétermination devient un problème quand on s'intéresse à l'usage *explicatif* de la théorie du choix rationnel. L'idée serait alors la suivante : supposons que l'ensemble des options envisageables soit A. Et supposons que plusieurs options, disons a_1 et a_2, sont compatibles avec la théorie du choix rationnel. Cela peut être, par exemple, tout simplement parce que l'agent est *indifférent* entre a_1 et a_2. Et supposons enfin que la personne qui nous intéresse ait choisi a_1. Dans ce cas, le problème serait peut-être que la théorie ne permet pas vraiment d'expliquer le choix de cette option puisque l'option a_2 était également envisageable du point de vue de la théorie. On pourrait d'ailleurs faire des remarques analogues pour la prédiction : la théorie ne permet pas de prédire l'option effectivement choisie mais un sous-ensemble des options auquel l'option choisie appartient. Il me semble néanmoins que, dans les deux cas, on a affaire à quelque chose d'essentiellement graduel et dont le caractère problématique est *a priori* très variable. Génériquement, si A' est l'ensemble des options compatibles avec la théorie dans une certaine situation étant donné

certaines informations sur l'agent, alors la théorie est *plus ou moins* explicative ou prédictive selon la différence entre A' et A. Mais même si A' n'est pas réduite à un singleton (cas de prédiction unique ou d'explication univoque), la théorie peut être utile et informative.

(C2) La seconde question que je voudrais poser porte sur l'importance accordée à l'indétermination dans la formation des croyances rationnelles. Plaçons-nous pour le moment dans un cadre non-stratégique : l'agent face à la « Nature ». La théorie dominante de la décision, la théorie bayésienne qui remonte sous sa forme contemporaine aux *Foundations of Statistics* (1954) de L. Savage, impose très peu de contraintes aux croyances des agents. Ces contraintes sont des contraintes de *structure* qui exigent que les degrés de croyance forment une distribution de probabilité. A l'intérieur de ces limites, elle laisse ces degrés de croyance varier d'un agent à l'autre, à peu près comme elle laisse varier le contenu des désirs d'un agent à l'autre. On peut remettre en question l'idée que les individus *sont* dotés de croyances probabilistes (ce que fait d'ailleurs J. Elster sur la base de données de la psychologie expérimentale) ; on peut même remettre en question l'idée que des agents rationnels *doivent* avoir des croyances probabilistes. Mais la théorie standard ne me semble pas avoir l'ambition de prescrire quelles sont « les » croyances rationnelles. Et *a fortiori* elle ne présuppose pas des agents qu'ils aient « ces » croyances rationnelles. Notons d'ailleurs que cela n'empêche pas la théorie de faire des prédictions : certains choix sont incompatibles avec la théorie (voir par exemple le fameux « Paradoxe d'Ellsberg ») *quelles que soient les croyances des agents*. Il me semble par conséquent que les remarques sur l' « incertitude brute » de Jon Elster portent non pas sur la théorie du choix rationnel, mais sur ce qu'on pourrait appeler des *programmes forts de logique inductive* : des programmes qui reposent sur l'idée que les croyances probabilistes d'un agent pourraient être déterminées univoquement par un certain nombre de règles bien définies. Le principe d'indifférence (ou principe de Laplace) selon lequel, par défaut, on attribuerait aux différentes possibilités envisageables une probabilité égale, pourrait, par exemple, être une telle règle. J'accepte tout à fait les critiques de J. Elster vis-à-vis du principe d'indifférence. Je reconnais également que de tels principes sont parfois utilisés en théorie du choix rationnel (le fameux argument bayésien dit de l'observateur impartial en faveur de l'utilitarisme élaboré par J. Harsanyi repose sur le principe d'indifférence). Mais l'articulation entre, d'une part, la théorie bayésienne de la décision, et, d'autre part, les questions que je viens de soulever, me semble délicate.

(C3) Les choses sont différentes avec la théorie des jeux, et j'en viens à mon troisième commentaire. Je laisse de côté la critique que J. Elster fait des équilibres en stratégie mixte – j'ai tendance à partager l'analyse de N. Houy -, pour me concentrer sur les jeux (et ils sont nombreux) où une multiplicité d'équilibres de Nash existent. C'est le cas des jeux comme celui de la Poule Mouillée ou de la Lutte des Sexes. L'équilibre de Nash ne permet pas de réduire à un seul les profils d'actions possibles. Dans ces jeux il y a *deux* équilibres symétriques. On pourrait considérer qu'il s'agit d'une indétermination au sens où nous l'avons vu précédemment : l'espace des issues possibles est l'ensemble des profils d'action, et la « solution » standard du jeu ne permet pas de réduire à un singleton cet ensemble. L'analyse de J. Elster suggère, à juste titre je crois, que la situation est plus problématique : le problème n'est pas seulement que dans ces jeux on ne voit pas pourquoi tel équilibre plutôt que tel autre serait joué, mais plutôt qu'on ne voit pas du tout pourquoi un des équilibres serait joué. Il semble que la multiplicité des équilibres ruinent la confiance qu'on peut avoir dans la notion-même d'équilibre. En effet, *si les joueurs ne peuvent pas se coordonner sur un équilibre précis, on ne voit pas pourquoi ils se coordonneraient sur un équilibre tout court*. La raison en est que, dans ces jeux, les équilibres ne sont pas « interchangeables » comme on dit parfois : si (s_1, s_2) et (s_1', s_2') sont des équilibres, il n'est pas vrai pour autant que (s_1,s_2') ou (s_1', s_2) le soient. Par conséquent, même si les deux joueurs jouent une stratégie figurant dans

un équilibre, il n'en découle pas que leurs stratégies forment conjointement un équilibre. Notons d'ailleurs que dans les expériences sur la Lutte des Sexes rapportées par C. Camerer, *Behavioral Game Theory* (2003), environ 60% des parties voient des profils « hors équilibre » se réaliser. Si je rassemble cette remarque et la remarque précédente (C2), alors j'ai le sentiment qu'il y a une asymétrie entre d'une part, l'importance de l'incertitude « brute » pour la théorie bayésienne de la décision et, d'autre part, l'importance de l'incertitude « stratégique » pour la théorie des jeux. Le second cas me paraît bien plus problématique que le premier pour la théorie. A vrai dire, ce n'est pas tellement surprenant puisque, avec la notion d'équilibre, la théorie des jeux, à la différence de la théorie de la décision individuelle, semble faire une hypothèse sur les croyances qu'un agent doit avoir.

(C4) Mon quatrième commentaire prolonge le troisième. L'interprétation des équilibres de la théorie des jeux comme l'équilibre de Nash a suscité une littérature formelle et philosophique immense. Et en premier lieu les analyses qui portant sur les « fondements épistémiques des équilibres ». Il me semble - je me trompe peut-être - que l'équilibre de Nash s'interprète de moins en moins, dans des jeux non-répétés, comme étant *la* théorie du choix rationnel. Dès lors, on peut se demander s'il est correct de faire de l'équilibre de Nash le pendant « stratégique » de la théorie de la décision individuelle. Peut-être, par exemple, l'élimination itérée des stratégies dominées constitue-t-elle une extension plus plausible. Cette solution retient en effet les stratégies qui ne sont pas dominées par d'autres stratégies dans le jeu initial, qui ne sont pas non plus dominées par d'autres stratégies dans le jeu formé à partir du jeu initial par élimination des stratégies dominées, et ainsi de suite. On peut d'ailleurs montrer que la solution découle de l'hypothèse selon laquelle les joueurs sont rationnels et partagent une croyance commune dans la rationalité les uns des autres. Ce qui est fort différent de supposer que les joueurs connaissent les actions des autres, comme c'est le cas avec l'analyse épistémique la plus simple de l'équilibre de Nash. Voici comment David Kreps, un éminent théoricien des jeux, décrivait récemment la situation :

> « Lorsqu'on applique le critère de dominance simple, on fait l'hypothèse implicite que les individus ne choisissent pas des stratégies dominées; quand on élimine celles-ci par itération, on fait l'hypothèse implicite que chaque joueur agit en supposant que les autres ne choisissent pas leurs stratégies dominées, et ainsi de suite. Pour autant que ces hypothèses soient correctes, le critère consistant à éliminer les stratégies dominées - y compris par itération - fournit un moyen très clair et direct de faire des prédictions.
>
> Avec l'équilibre de Nash, la « logique » est beaucoup moins claire. Il est vrai que dans certains cas, chaque participant voit de façon assez évidente quel doit être son choix et celui des autres. Dans de tels cas, les choix « évidents » qui s'imposent ainsi à tous constituent nécessairement un équilibre de Nash...A moins qu'un jeu ait une façon de jouer qui semble évidente, il n'y a pas de raison d'accorder une place particulière à la notion d'équilibre de Nash. […] les grandes difficultés que rencontre la théorie des jeux proviennent de ce que l'on ne voit pas clairement (pour ne pas dire plus) quand et pourquoi l'analyse par équilibre est pertinente, si elle l'est. »[1]

Avec l'élimination itérée des stratégies dominées, le caractère « spécialement dramatique » de l'indétermination dans un contexte stratégique disparaît : le fait que plusieurs profils d'actions soient compatibles avec cette solution ne constitue pas une raison particulière

[1] D. Kreps, *Game Theory and Economic Modelling*, Oxford University Press, 1990 ; trad.fr. *Théorie des jeux et modélisation économique*, Dunod, 1999.

de se défier du fait que les profils effectivement joués seront compatibles avec la solution. En revanche, je reconnais que, en ce qui concerne l'élimination itérée des stratégies dominées, le problème « ordinaire » de l'indétermination peut prendre un tour préoccupant : l'élimination itérée peut ne rien éliminer du tout et donc ne fournir aucune information prédictive ou explicative.