Sur la rhétorique de la théorie des jeux

(Chapitre 5 de *Economics and language* de Ariel Rubinstein)

Ariel Rubinstein

1. Introduction

Ce chapitre contient quelques commentaires sur le langage utilisé par les théoriciens des jeux. Il est donc différent du reste de cet ouvrage. Pourquoi me suis-je intéressé à la rhétorique de la théorie des jeux ? En fait, je ne suis pas enthousiasmé par le fait que la théorie des jeux soit vue par beaucoup comme "utile" au sens où elle procurerait un guide pour agir dans des situations stratégiques, vision renforcée par le langage utilisé par les théoriciens des jeux.

Considérons, par exemple, l'ouvrage de John Mc Millan, *Games, Strategies and Managers* (1992). Sur la quatrième de couverture de l'édition brochée, on cite Fortune appréciant l'ouvrage comme « le guide le plus sympathique pour les hommes d'affaires » et Akira Omori de Arthur Andersen disant que l'ouvrage « sera utile à la fois pour les managers débutants et pour les planificateurs de stratégie expérimentés. En fait je voudrais recommander à quiconque est engagé dans les négociations USA-Japon de lire cet ouvrage ».

Considérons encore le best-seller de Avinash Dixit et Barry Nalebuff, *Thinking Strategically*. Le *Financial Times* du 7 décembre 1991 dit de l'ouvrage : « *Thinking Strategically*... offre un apprentissage essentiel dans les choix et les possibilités d'agir non seulement dans les affaires mais aussi dans la vie quotidienne. » Shlomo Maital dit, dans sa revue de l'ouvrage, publiée dans *Across the Board* (juin 1991) : « Jamais depuis 1926, lorsque John von Neumann, un brillant mathématicien et physicien hongrois, publia son papier novateur sur la théorie des jeux, on avait perçu que des modèles analytiques des jeux pouvaient être construits. Ces modèles permettent de trouver les meilleures stratégies et les meilleures solutions pour les jeux, y compris pour ceux qui comportent des dilemmes d'affaires complexes. »

La fausse conception de ces livres est renforcée par les auteurs eux-mêmes. John Mc Millan dit : « Dépouillé de ses mathématiques et de son jargon, il peut être utile pour les personnes qui se trouvent dans des situations de management comportant des décisions stratégiques importantes », et la phrase ouvrant le livre indique : « Qu'en est-il de la théorie des jeux et comment peut-elle être utilisée dans la prise de décision ? » Dixit et Nalebuff écrivent : « Il est préférable d'être un bon stratège qu'un mauvais et ce livre cherche à aider à améliorer votre aptitude à découvrir et utiliser des stratégies effectives. » « Notre objectif est d'améliorer votre QI stratégique » et « nos présupposés, en écrivant ce livre, sont que les lecteurs deviennent de meilleurs stratèges s'ils connaissent ces principes ».

Le succès phénoménal de la théorie des jeux et la perception selon laquelle ils peuvent « améliorer le QI stratégique » sont liés. Je doute que la théorie des jeux aurait attiré l'attention à ce point si elle avait été appelée « la théorie des interactions entre des agents rationnels ». Le fait que l'on puisse trouver un livre "à propos"» de la théorie des jeux dans la liste des best-sellers du *New York Times* et que les étudiants soient fascinés par le « dilemme du prisonnier » provient des associations naturelles du jargon de la théorie des jeux : des mots tels que "jeu", "stratégie" et "solution" stimulent notre imagination. L'utilisation de ces mots est-elle justifiée ?

Les mots sont une partie cruciale des modèles économiques. Un modèle économique diffère substantiellement d'un modèle purement mathématique parce qu'il est une *combinaison* d'une structure mathématique et d'une *interprétation.* Les noms des objets mathématiques font partie *intégrante* du modèle économique. Lorsque les mathématiciens utilisent des mots courants tels que "groupes" ou "anneau", c'est seulement par commodité. Lorsqu'ils nomment une collection d'objets un "filtre", ils le font par association et, en principe, ils pourraient l'appeler aussi bien "cornet de glace". Lorsqu'ils utilisent le terme "bon ordre", ils n'y assignent aucune valeur éthique positive. La situation est quelque peu différente en logique mathématique dans laquelle, par exemple, les noms des connecteurs sont fortement reliés à leur signification commune. Dans la théorie économique, et dans la théorie des jeux en particulier, l'interprétation est un ingrédient essentiel de tout modèle. Un jeu varie suivant que les joueurs sont des êtres humains, des abeilles ou différents "moi" de la même personne. Un jeu stratégique change entièrement si les gains passent des valeurs d'utilité représentant les préférences de von Neumann et Morgenstern aux mesures de pertinence évolutionnaire *(evolutionary fitness).*

L'évaluation de la rhétorique de la théorie des jeux est aussi importante du point de vue du public. L'intérêt du public pour la théorie des jeux est, au moins partiellement, le résultat des efforts des économistes théoriciens qui mettent l'accent sur la valeur pratique de la théorie des jeux comme, par exemple, guide pour les décideurs politiques. Des consultants utilisent le langage professionnel de la théorie des jeux dans leurs argumentations, et il est dans l'intérêt du public que ces arguments soient compris correctement de sorte que la rhétorique ne déforme pas leur contenu réel. Quoique la discussion de l'utilité de la théorie des jeux se situe à un haut niveau d'abstraction, sans se référer à la substance des modèles théoriques de jeu, elle requiert une compréhension de la théorie des jeux qui va au-delà de celle requise pour le "dilemme du prisonnier".

Sur la rhétorique de la théorie des jeux

J'ai des doutes sur l'applicabilité pratique de la théorie des jeux. Cependant, je ne me sens pas pessimiste parce que je ne considère pas l'applicabilité comme une vertu indispensable. Je ne crois pas que l'étude de la logique formelle puisse aider les gens à devenir "plus logiques", et il ne me paraît pas évident que l'étude de la théorie des probabilités améliore significativement la capacité des individus à raisonner en termes probabilistes. La théorie des jeux est reliée à la logique et à la théorie des probabilités, et je doute qu'elle puisse s'avérer utile aux négociateurs ou à d'autres types de joueurs. En fait, je me demande si la théorie des jeux ne pourrait pas, au contraire, induire les gens en erreur du fait que beaucoup ignorent les subtilités des argumentations théoriques sur les jeux et traitent les solutions comme des instructions. (Néanmoins, je me dois de mentionner que je ne suis pas parvenu à démontrer cela dans une expérience menée en collaboration avec un groupe d'étudiants de l'université de Tel Aviv. Les résultats de l'expérience n'ont pas permis de conclure.)

En fait, l'ouvrage de Mc Millan est plein de mises en garde contre l'utilité directe de la théorie des jeux : « La théorie des jeux ne doit pas prétendre dire aux managers comment faire tourner leur affaire » ; « la théorie des jeux n'élimine pas le besoin de la connaissance et de l'intuition acquises par une longue expérience » ; « la théorie des jeux offre un raccourci pour comprendre les principes de la prise de décision. Les managers qualifiés et expérimentés comprennent ces principes intuitivement, mais pas nécessairement d'une manière qui leur permette de communiquer leur compréhension aux autres » ; et « la théorie des jeux, donc, est une aide limitée mais puissante pour comprendre les interactions stratégiques ». Dixit et Nalebuff (en dépit de leur déclaration : « N'entrez pas en compétition sans elle ») écrivent que « d'une certaine manière, penser stratégiquement demeure un art ».

Et, en effet, lorsqu'il s'agit de décrire l'utilité de la théorie des jeux dans des cas réels, tous les auteurs adoptent des objectifs modestes et deviennent plutôt vagues. En dépit de la promesse d'être un guide sympathique pour l'utilisateur, Mc Millan conclut son chapitre "Negociating" – un sujet qu'il considère comme le problème archétypal de la théorie des jeux – par cette déclaration modeste : « Quel conseil pour les négociateurs la théorie des jeux peut-elle produire ? L'idée la plus importante que nous avons apprise dans ce chapitre est celle de la valeur de se mettre soi-même à la place des autres et de regarder plusieurs coups à l'avance. » La "règle n° 1" de Dixit et Nalebuff est « regarder en avant et raisonner en remontant en arrière ». Ils établissent le principe général pour les jeux séquentiels selon lequel « chaque joueur devrait imaginer les réponses futures des autres joueurs et à partir de celles-ci calculer sa meilleure action possible ». La phrase finale de l'ouvrage indique : « Vous pouvez parier que les machines où l'on mise le plus ne sont pas celles qui ont le plus haut rendement. »

Si l'enseignement le plus important à retirer de la lecture d'un « guide sympatique pour l'utilisateur de la théorie des jeux » est de se mettre à la place des autres personnes et si, l'amélioration de votre QI stratégique est conçue pour « anticiper la réponse de votre rival », je ne suis pas convaincu que la théorie des jeux soit plus valable pour atteindre cet objectif qu'un roman policier, un poème romantique ou un jeu d'échec.

D'ailleurs est-il clair que la connaissance académique soit nécessairement évaluable ? Est-il clair que l'amélioration du QI stratégique soit un but à rechercher ? David Walsh, un collaborateur du *Washington Post* a écrit : « Le problème est, évidemment, que si Dixit et Nalebuff peuvent améliorer votre QI stratégique, ils peuvent aussi bien améliorer celui de votre concurrent. » Ceci soulève une question importante qui vaut aussi pour les autres sciences sociales : améliorer le raisonnement stratégique, à supposer que ce soit possible (et j'en doute), devrait être évalué selon les standards sociaux qui prennent en compte non seulement les succès de la stratégie d'une entreprise mais aussi le bien-être social. Est-il souhaitable d'améliorer le QI stratégique d'un petit secteur de la population, spécialement un de ceux dont on ne souhaite pas voir améliorer la position ?

Incidemment, la discussion de la rhétorique de l'économie n'est pas un sujet nouveau. La recherche la plus connue sur le sujet a été publiée dans le milieu des années quatre-vingts (Mc Closkey, 1998, 2e édition) et a été suivie d'une abondante littérature (voir, par exemple, Henderson, Dudley-Evans et Backhouse, 1993). Ma contribution dans ce chapitre est assez limitée. Je voudrais mettre l'accent sur quelques problèmes rhétoriques associés au langage de la théorie des jeux. Je voudrais montrer que la rhétorique de la théorie des jeux est actuellement trompeuse en ce sens qu'elle laisse l'impression qu'elle a plus d'applications qu'elle n'en a en réalité. En particulier je veux montrer que :

1. les gens consultent la théorie des jeux pour savoir quelle stratégie adopter dans des situations analogues à un jeu ; pourtant, la notion de base de stratégie dans la théorie des jeux peut difficilement être interprétée comme une ligne d'action ;

2. l'utilisation de formules dans la théorie des jeux crée une illusion de précision qui n'a pas de fondement dans la réalité.

2. Stratégie

Stratégie dans le langage quotidien

L'un des termes centraux dans la théorie des jeux est celui de "stratégie". Qu'est-ce qu'une stratégie dans le langage courant ? Mc Millan (1992) nous apprend que l'origine du mot "stratégie" est le mot grec désignant le chef d'une armée, qui est assez différent de son usage aujourd'hui. *Le dictionnaire de Webster* définit le mot comme « une méthode pour produire ou faire quelque chose ou atteindre une fin » et le *Oxford English*

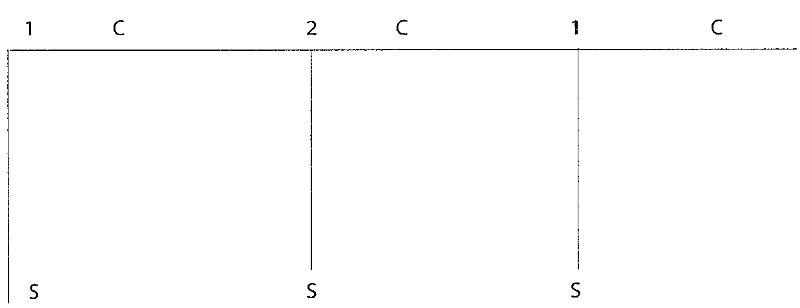
Dictionnary définit le mot comme « un plan général d'actions ». Nous parlons de stratégie pour gagner une guerre, de la stratégie des managers pour faire des profits ou éviter d'être licenciés et de la stratégie pour survivre à un ouragan.

La définition informelle d'une stratégie, selon les théoriciens des jeux, n'est pas éloignée de celle du langage quotidien. Martin Shubik fait référence à la stratégie comme « une description complète de la manière dont un joueur a l'intention de jouer un jeu, du début jusqu'à la fin ». Jim Friedman la définit comme « un ensemble d'instructions ». Et John McMillan définit une stratégie comme « une spécification des actions répondant à toutes les éventualités ». Ainsi, il est pertinent de poser la question de savoir si la définition formelle du terme "stratégie" dans la théorie des jeux est similaire à sa signification habituelle. Si elle ne l'est pas, alors on peut douter sérieusement de l'utilité de la théorie des jeux ; et même si elle l'est, il y a encore un long chemin à parcourir avant que nous soyons persuadés que les résultats de la théorie des jeux puissent apporter quelque chose aux managers.

Une stratégie dans un jeu extensif

Dans un jeu extensif, une stratégie de joueur est nécessaire pour spécifier une action à chaque nœud de l'arbre du jeu sur lequel le joueur doit se déplacer. Par conséquent, un joueur doit spécifier une action pour toute séquence d'événements qui est cohérente avec les règles du jeu. Comme illustration, considérons le jeu à deux joueurs de la figure 1.

Figure 1



Selon la définition naturelle d'une stratégie comme « plan d'action complet », le joueur 1 doit spécifier son comportement, "continuer" (C) ou "stopper" (S) au nœud initial et s'il prévoit de continuer de faire des plans prévisionnels pour le deuxième nœud de décision au cas où le joueur 2 décide de continuer. En revanche, la définition de la stratégie selon la théorie des jeux suppose que le joueur 1 spécifie son action au second nœud de décision, même si il prévoit de "stopper" le jeu au premier nœud. Ici, comme dans tout jeu où un joueur doit effectuer au moins deux mouvements consécutifs (et un grand nombre de jeux analysés dans la théorie économique entrent dans cette catégorie), une stratégie doit spécifier les actions des joueurs même après des événements qui sont incompatibles avec sa propre stratégie.

Pourquoi la notion de stratégie utilisée par les théoriciens des jeux diffère-t-elle de celle de "plan d'action" ? Si nous n'étudions que les équilibres de Nash dans les jeux extensifs, il ne serait en effet pas nécessaire de se doter d'une définition aussi large que celle de la théorie des jeux. La définition large est cependant nécessaire pour tester la rationalité des plans des joueurs aussi bien au début du jeu qu'au point où ils sont obligés d'envisager la possibilité de répondre à une déviation potentielle de l'adversaire (l'idée du sous-jeu parfait). En revenant à la forme du jeu ci-dessus, supposons que chaque joueur choisisse de "stopper" au premier nœud de décision. Tester l'optimalité du plan du joueur 2 à la suite de la déviation du joueur 1 suppose que le joueur 2 spécifie ses anticipations du plan du joueur 1 à son second nœud de décision. La spécification de l'action du joueur 1, après que les deux joueurs aient choisi "stopper", détermine ces anticipations et doit être interprétée comme ce qu'aurait été la croyance *(belief)* du joueur 2 (en réponse au joueur 1) quant à ce que prévoit de jouer le joueur 1 qui décide de dévier de ce qu'il croyait être son plan d'action originel. Par conséquent, une stratégie n'intègre pas seulement le plan du joueur mais aussi les anticipations de ses adversaires au cas où il ne suivrait pas ce plan. Par conséquent, une stratégie d'équilibre décrit un plan d'action du joueur aussi bien que les considérations qui expliquent l'optimalité de ce plan (c'est-à-dire les idées préconçues concernant les plans des autres joueurs). Ce n'est pas seulement une description d'"un plan d'action". Un profil de stratégies offre une analyse complète de la situation et pas seulement un n-uple des plans d'actions.

Une stratégie mixte dans un jeu stratégique

L'interprétation naïve de l'expression "stratégie mixte" se réfère à un joueur qui utilise une roulette ou un autre moyen aléatoire pour décider d'une action. En général, nous sommes réticents pour admettre que nos décisions soient prises au hasard ; nous préférons croire qu'il y a une raison pour chacune des actions que nous entreprenons. Un grand nombre de théoriciens des jeux n'ont jamais envisagé l'utilisation de moyens aléatoires comme partie intégrante d'une stratégie avant d'être confrontés à la théorie des jeux. En effet, le concept de stratégie mixte a été l'objet de critiques. Pour citer Aumann (1987) : « Les équilibres de stratégie mixte ont toujours été intuitivement problématiques » et Radner et Rosenthal (1982) : « L'une des raisons pour lesquelles la théorie des jeux n'a pas trouvé d'application plus étendue est que la randomisation, qui joue un rôle majeur en théorie des jeux, semble avoir un intérêt limité dans beaucoup de situations pratiques. »

Il y a évidemment des cas dans lesquels les joueurs choisissent des actions aléatoires. Lorsqu'un principal (par exemple, un employeur ou le gouvernement) contrôle ses agents (employés ou citoyens), c'est souvent au hasard qu'un petit nombre d'entre eux sont choisis pour être contrôlés. Evidemment, il ne s'agit pas d'une "stratégie mixte". La règle stochastique utilisée pour mener les enquêtes est en fait une stratégie pure. Le principal n'est pas indifférent entre contrôler aucun agent et les contrôler tous. Sa "stratégie pure" est le choix des paramètres de la randomisation à utiliser. Autre exemple : un père, ou une mère, qui doit donner un paquet de bonbons à l'un de ses deux enfants peut préférer désigner le bénéficiaire en jouant à pile ou face. Cependant, jouer à pile ou face n'est pas non plus une "stratégie mixte". En effet les parents ne sont pas indifférents à la procédure choisie pour attribuer le paquet de bonbons. Une procédure d'allocation peut être considérée comme une stratégie pure, même si elle comporte un jeu de pile ou face.

La littérature sur la théorie des jeux suggère quelques interprétations supplémentaires de ce concept. L'interprétation d'une stratégie mixte en termes de fréquence dans une population nombreuse a été utilisée dans le chapitre 2 de cet ouvrage. Le jeu est parfois joué parmi les joueurs, chacun étant tiré au sort dans une population homogène dans ses intérêts mais hétérogènes dans ses actions. La stratégie mixte décrit la distribution des actions choisies par les agents issus d'une population attachée à un rôle donné dans le jeu. Dans une interprétation différente, une stratégie mixte est conçue comme un plan d'action qui est dépendant d'une information privée non spécifiée dans le modèle. Selon cette interprétation, le comportement d'un joueur est en fait déterminé bien qu'il apparaisse comme aléatoire. Si nous ajoutons cette structure d'information au modèle, la stratégie mixte devient une stratégie pure dans laquelle les actions dépendent d'une information externe. Cette interprétation est problématique parce qu'on ne voit pas clairement pourquoi les individus pourraient baser leur comportement sur des facteurs qui sont clairement sans rapport avec la situation (voir Harsanyi, 1973, pour une interprétation plus sophistiquée qui répond à cette critique).

Il y a encore une autre interprétation qui a gagné en popularité au cours de la dernière décade. Selon celle-ci, une stratégie mixte est la croyance, concernant les actions d'un joueur particulier, adoptée par tous les *autres* joueurs (voir Aumann, 1987). Un équilibre en stratégie mixte est alors un n-uple des anticipations de connaissance commune ; il a la propriété que, pour des croyances données, toutes les actions ayant une probabilité strictement positive sont optimales. Il en résulte que les incertitudes qui sous-tendent l'équilibre en stratégie mixte sont vues comme une expression du manque de certitude de la part des autres joueurs, plutôt que du plan délibéré d'un joueur individuel.

Quelle importance ?

Il apparaîtrait donc que deux versions centrales de la notion de "stratégie" dans la théorie des jeux – la stratégie dans la forme extensive et la stratégie mixte dans les jeux stratégiques – sont interprétées de manière plus appropriée en termes de croyance qu'en termes de plans d'actions. Cependant, si les stratégies sont des croyances plutôt que des plans d'actions, toute la façon de présenter la théorie des jeux doit être revue. Nous pouvons parler du joueur 1 choisissant sa stratégie, mais peut-il choisir la croyance du joueur 2 à propos de sa stratégie ?

Comme démonstration des difficultés entraînées par l'interprétation des croyances, considérons la littérature sur les jeux séquentiels, dans laquelle les auteurs *supposent* que les stratégies sont stationnaires, en ce sens que le comportement d'un joueur est indépendant de l'histoire du jeu. La littérature présente la stationnarité comme une hypothèse relative à la simplicité du comportement. Par exemple, considérons la stratégie du joueur 1 "toujours jouer *a*" dans un jeu répété. Cette stratégie est simple en ce sens que le joueur prévoit de jouer indépendamment des actions des autres joueurs. Mais cette stratégie suppose également que le joueur 2 croit que le joueur 1 jouera *a* même si le joueur 1 a joué *b* dans les 17 premières périodes du jeu. La stationnarité dans les jeux séquentiels n'implique donc pas seulement la simplicité mais aussi la passivité des croyances. Ceci est contre-intuitif, surtout si nous faisons l'hypothèse de la simplicité des comportements. Si le joueur 2 croît que le joueur 1 est contraint de choisir un plan d'actions stationnaire, alors le joueur 2 devrait croire probablement (après 17 répétitions de l'action *b*) que le joueur 1 continuera de jouer *b*. Ainsi, en faisant l'hypothèse de la passivité des croyances, nous éliminons un important problème que les jeux séquentiels cherchent à modéliser : structure changeante du comportement du joueur, croyances variant avec l'expérience accumulée du joueur.

Considérons maintenant la question des stratégies mixtes. En admettant l'interprétation par les croyances, nous sommes conduits à une réévaluation profonde des applications de la théorie des jeux. En particulier, cette interprétation implique que l'équilibre ne permet pas une prédiction statistique du comportement des joueurs. En effet, toute action entreprise par un joueur *i* qui est une meilleure réponse à une anticipation donnée des stratégies des autres joueurs, est une prédiction cohérente de la future action de *i* (ceci peut même inclure les actions qui sont à l'extérieur du support de la stratégie mixte). Ceci prive de signification toute statique comparative ou analyse du bien-être de l'équilibre en stratégie mixte et met en question l'énorme littérature qui utilise l'équilibre en stratégie mixte.

3. L'illusion du "nombre"

Bien que la question de savoir si l'on doit représenter un individu par une relation de préférence ou par une fonction d'utilité puisse sembler sans portée concrète (puisqu'une fonction d'utilité n'est qu'une représentation numérique d'une relation de préférence), je voudrais soutenir que ce n'est pas le cas et que la « numérisation de la théorie des jeux » envoie un message trompeur quant à la précision et à la pertinence de la théorie des jeux.

Pour démontrer ce point, discutons rapidement la solution de négociation de Nash, qui est l'un des modèles les plus fondamentaux de la théorie économique moderne. La théorie de Nash est conçue pour fournir une "prédiction" du résultat de la négociation, basée sur deux éléments :

1. les préférences du négociateur, définies sur l'ensemble des accords possibles (y compris le cas de désaccord) ;
2. l'attitude du négociateur face au risque.

Les primitives du problème de la négociation de Nash (à deux personnes) sont *l'ensemble faisable* et *le point de désaccord,* notés respectivement S et d. Dans la formalisation de Nash du problème de la négociation, chaque élément de S correspond à une paire de nombres interprétés comme les niveaux d'utilité obtenus par les deux négociateurs pour l'un (ou plus) des accords possibles. L'interprétation des utilités est celle de Von Neumann-Morgenstern en ce sens qu'elles sont dérivées de relations de préférence entre les loteries qui satisfont les hypothèses de l'utilité espérée. Le point de désaccord est modélisé comme un point de S. Le second concept de base de la théorie de la négociation de Nash est la *solution de la négociation,* définie comme une fonction qui attribue une paire unique de niveaux d'utilité à chaque problème de négociation (S, d). Ainsi, une solution de négociation est supposée fournir une "prédiction unique" au résultat de la négociation (en termes d'utilité) pour chacun des problèmes dans son domaine.

Nash (1950) a montré qu'il y a une unique solution de négociation satisfaisant les quatre axiomes suivants : *invariance à une transformation affine positive* (le changement d'échelle des valeurs d'utilité de chaque négociateur par une transformation positive affine déplace la solution de la même façon) ; *symétrie* (un problème symétrique a un point de solution symétrique) ; *optimalité de Pareto* (pour chaque problème de négociation, la solution est sur la frontière de Pareto) ; et *indépendance par rapport aux choix non pertinents* [IIA] (1) (la restriction de l'ensemble des alternatives sans élimination du point de solution ni du point de désaccord ne modifie pas le point de solution), ce qui est le plus problématique des axiomes en termes d'interprétation. La solution unique satisfaisant ces axiomes est la solution de Nash, c'est-à-dire la fonction :

$N(S,d) = \text{argmax } \{(u_1 - d_1)\ (u_2 - d_2)\}\ (u_1, u_2) \in S$ et $u_i \geq d_i$ pour les deux valeurs de i

(1) [Independance of irrelevant alternatives.]

Si l'objet de la théorie de la négociation est de fournir une prévision quantitative "claire et nette" pour un grand nombre de problèmes de négociation, il est évident que Nash a atteint cet objectif. Le fait que cette solution soit définie par une formule simple est un avantage significatif de la théorie, spécialement lorsqu'on essaie de l'insérer dans un modèle plus vaste qui contient une composante de négociation. Mais cette prévision peut-elle être testée comme dans les sciences ? Cette formule prédit-elle comment les gens vont partager un gâteau avec la même exactitude que nous pouvons calculer le moment où une pierre, tombant d'une tour, va atteindre le sol ?

J'en doute très fortement. L'importance de la formule de Nash vient de sa signification abstraite, indépendamment du fait qu'elle soit vérifiable ou non. L'utilisation de nombres, même si elle est analytiquement pertinente, obscurcit le sens du modèle et crée l'illusion qu'il peut produire des résultats quantitatifs. La solution de Nash ne devrait-elle pas avoir une signification différente, plus abstraite ?

Face à cette question, la réponse est "non". La formule de la solution de Nash à la négociation n'a pas une signification claire. Quelle est, en effet, l'interprétation du produit des nombres représentant l'utilité de von Neumann-Morgenstern ? Quel est le sens de la maximisation de ce produit ? Pouvons-nous considérer la maximisation du produit des utilités comme un principe "utile" pour résoudre les conflits ?

L'usage de nombres pour spécifier le problème de la négociation a obscurci le sens de la solution de Nash. Si les théoriciens des jeux avaient adopté un langage plus naturel pour spécifier le modèle, la solution serait devenue plus claire et plus significative. Dans Rubinstein, Safra et Thomson (1992) le langage indirect de l'utilité est remplacé par le langage direct des préférences. Les éléments du modèle sont interprétés directement en utilisant les mots "choix", "loterie", "désaccord", "préférence". Le problème de Nash est remplacé par un quadruplet $(X, D, \geq_1, \geq_2)$ où X est un ensemble d'accords réalisables, D est un événement de désaccord et $(\geq_1, \geq_2)$ sont les relations de préférence des négociateurs.

Comme la théorie vise à prédire le résultat d'une négociation en fonction des intérêts des joueurs et de leur attitude face au risque, les préférences sont définies sur l'ensemble des loteries dont les accords dans X et les événements de désaccord D sont "des lots certains".

En utilisant ce modèle, une définition alternative de la solution de Nash peut être formulée :

Une solution (ordinale) de Nash pour le problème $(X, D, \geq_1, \geq_2)$ est un choix y^* tel que pour tout $p \in [0,1]$ et tout $x \in X$ et tout i, si $p^{\circ}x \geq_i y^*$ alors $p^{\circ}y^* \geq_j x$ (où $p^{\circ}x$ est une loterie qui donne x avec la probabilité p et D avec la probabilité $1 - p$).

L'interprétation suggérée de cette solution est la suivante : une solution est une "convention" qui fait correspondre un unique accord à tout problème de négociation. La solution comporte l'hypothèse que les joueurs sont conscients du fait que s'ils soulèvent une objection à un choix, ils courent le risque que les négociations se terminent par un désaccord. L'accord qui est solution de Nash est le seul qui satisfasse la propriété suivante :

si

- il est intéressant pour l'un des joueurs de demander une amélioration dans la convention, même en risquant une rupture des négociations,

alors

- il est optimal pour l'autre joueur d'insister pour l'application de cette convention, même au risque d'une rupture des négociations.

En d'autres termes, la solution de Nash à la négociation est un accord satisfaisant à condition qu'un argument du type « vous devriez accepter ma demande x sans délai puisqu'elle est préférable pour vous à la convention y* compte tenu de la probabilité de rupture p » n'est pas profitable pour le joueur faisant cette demande lorsqu'il prend en compte la même probabilité de rupture.

La guerre du Golfe fournit un exemple concret de cette définition : les négociateurs étaient l'Irak et les USA. L'ensemble des accords contient l'ensemble des différentes partitions possibles du territoire de la région. L'événement de désaccord est la guerre. Lorsque Sadam Hussein déplaça ses troupes, il tenta la chance que la situation dégénère en guerre avant que les USA capitulent. Apparemment, il préféra le risque de la guerre avec une certaine probabilité parce qu'il espérait que les USA accepteraient sa demande d'annexion du Koweit. Contrairement à ce qu'il avait anticipé, les USA ont préféré prendre le risque de la guerre et demander un retour au *statu quo* plutôt que d'accéder à la demande irakienne. Si les USA avaient cédé à l'Irak, cela aurait signifié que les frontières antérieures à l'invasion n'étaient pas incluses dans le résultat de la négociation de Nash.

Lorsque les préférences satisfont les hypothèses de l'utilité espérée, cette définition et la définition standard coïncident (voir Rubinstein, Safra, Thomson, 1992). Cependant, un produit joint de la définition ordinale est que la théorie de Nash peut être étendue à un domaine de préférences plus large que celui de la théorie de l'utilité espérée. Le langage des préférences ne nécessite pas que les préférences satisfassent les axiomes de l'utilité espérée ; il en résulte que la définition peut être appliquée à un ensemble de préférences sur l'ensemble des loteries, portant sur les accords et le désaccords, plus vaste. On peut montrer que, pour une classe plus large de préférences (qui inclut toutes les préférences satisfaisant les axiomes de l'utilité espérée), la solution de Nash est bien définie et est la seule solution qui satisfasse un ensemble d'axiomes qui ressemblent aux axiomes de Pareto, de symétrie et d'indépendance par rapport aux choix non-pertinents (IIA).

En fait, le passage au langage des préférences sur les choix implique une refonte de la théorie de Nash toute entière. Les détails de cette démarche n'entrant pas dans les objectifs de ce livre, je vais seulement commenter les changements dans l'axiome IIA. L'axiome IIA d'origine exige que si u^* est paire d'utilité solution du problème (T, d) et si u^* est un élément d'un ensemble S qui est un sous-ensemble de T, alors u^* est aussi solution du problème (S, d). Comme on l'a souvent souligné, la justification de cet axiome correspond à une théorie normative dans laquelle le concept de solution est vu comme le reflet de la désirabilité sociale d'un choix. Lorsque la négociation est vue comme une interaction stratégique entre des négociateurs qui ne se soucient que de leur propre intérêt, on peut s'interroger sur la validité de cet axiome IIA.

Ce qui suit est un traitement différent de l'axiome IIA qui ne requiert pas de comparaison entre des problèmes qui ont différents ensembles de choix. Le "IIA ordinal" pose que la solution y^* d'un problème de négociation demeure invariante lors d'un changement de préférence du joueur i qui abaisserait sa volonté de soulever des objections à la solution y^*. Formellement, supposons que y^* est une solution du problème $(X, D, \geq_1, \geq_2)$, et soit $\geq'_i$ une relation de préférence qui soit en accord avec $\geq_i$ sur l'ensemble des accords X tels que :

(i) pour tout x tel que $x \geq_i y^*$, si $p°x \sim_i y^*$ alors $p°x \leq_i y^*$ et

(ii) pour tout x tel que $\leq_i y^*$, si $x \sim'_i p°y^*$ alors $p°x \leq_i y^*$.

Il en résulte que y* est aussi la solution du problème :

$(X, D, (X, D, \geq_1, \geq_2), \geq'_2)$.

Le passage du joueur i de la relation de préférence $\geq_i$ à la relation de préférence $\geq'_i$ reflète une aversion accrue vis-à-vis du risque qu'il encourrait en demandant des choix qui seraient meilleurs pour lui que le résultat y^*. Le changement dans les préférences du joueur i le rend « moins désireux d'exprimer une objection » à l'accord. Bien que le joueur i ait, avec la relation $\geq'_i$ le même préordre des préférences qu'avec la relation $\geq_i$, il est moins volontaire pour prendre le risque de demander un accord qui soit meilleur que y^*. L'axiome exprime que le changement n'affecte pas le résultat de la négociation. En d'autres termes, changer les préférences du joueur i, en le rendant moins désireux d'atteindre un accord meilleur que y^*, ne doit pas changer le résultat de la négociation. Avec cette interprétation de l'axiome IIA, le lien entre l'axiomatisation et la définition ordinale de la solution de Nash semble plus compréhensible. L'axiome prend son sens lorsque la procédure sous-jacente à la négociation est telle qu'il y a une convention claire et que l'objection peut être "protégée" des autres objections seulement en insistant sur la conservation de la convention. Il a moins de sens lorsqu'un contre-argument peut être soutenu par d'autres moyens, comme des demandes alternatives croissantes, ou lorsqu'un accord est atteint par les deux parties en réduisant graduellement leurs demandes.

Revenons au point principal de cette section : le langage de l'utilité permet l'utilisation de présentations géométriques et facilite l'analyse ; en revanche, la présentation cardinale repose sur un traitement non naturel des axiomes de la solution. Le passage au langage des préférences permet un traitement plus naturel du traitement et de la caractérisation de la solution de Nash. On peut avancer que l'équivalence entre les termes "préférences" et "utilités" dans les modèles économiques implique aussi l'équivalence dans leur utilisation. Cependant, comme on l'a démontré ici, le choix du langage affecte la signification des résultats. L'utilisation du langage de l'utilité appelle à des opérations arithmétiques (telles que la multiplication des utilités) dont les interprétations n'ont pas nécessairement de signification. L'utilisation du langage des préférences conduit à une solution cohérente avec les considérations qui peuvent se concevoir dans les usages de la vie réelle. Néanmoins, je ne suis pas convaincu que la théorie de Nash ait fait plus que clarifier la logique d'une considération qui influence le résultat de la négociation. Je ne parviens pas à voir comment cette considération peut expliquer complètement les résultats de la négociation dans la vie réelle.

4. Le terme "solution"

Dans les deux sections précédentes, j'ai discuté l'usage du terme "stratégie" et la substitution du terme "utilité" par le terme "préférence". Dans cette section, je veux discuter brièvement un autre mot-clé de la théorie des jeux, celui de "solution". La structure analytique de la théorie des jeux est la suivante : premièrement, le modélisateur explicite les paramètres du jeu (identités des joueurs, règles, information disponible pour les joueurs et intérêts des joueurs). L'analyse consiste ensuite à « trouver des solutions ». Le résultat de l'application du concept de solution à un jeu est un ensemble de profils, chacun d'entre eux attribuant une stratégie à chacun des joueurs. Un concept de solution est une méthode qui assigne à chaque jeu un ensemble de profils de stratégies satisfaisant certaines conditions de stabilité et de rationalité.

Cette forme de recherche fait une distinction entre les hypothèses incluses dans le concept de solution et celles qui sous-tendent la description du jeu. Les premières sont considérées comme les plus solides, bien qu'elles soient en fait des hypothèses très fortes concernant le comportement des joueurs et le processus de délibération. Comme elles sont enfouies dans le concept de solution, ces hypothèses ne sont pratiquement jamais discutées par les utilisateurs de la théorie des jeux. Notons que ces hypothèses sont tout sauf claires même pour les théoriciens des jeux, et il a fallu un long délai avant que les théoriciens examinent les hypothèses épistémologiques qui sous-tendent les différents concepts de solution.

Il en résulte que, tandis que les économistes appliqués débattent souvent à propos des hypothèses du modèle, ils n'accordent guère d'attention à la pertinence des hypothèses qui sont intégrées dans le concept de solution. Un concept de solution est traité comme une machine dans laquelle on fait entrer la description de la situation et qui fournit un produit qui est interprété comme une prévision sur le moyen par lequel le jeu va être résolu.

A mon avis, l'utilisation du terme "solution" joue un rôle central pour donner cette vision de la théorie des jeux comme une « machine de prédiction ». Le terme « solution » implique l'existence d'un problème. Dans le cas de la théorie des jeux, le problème est d'arriver à comprendre « ce qui va se passer » ou « quelle décision prendre ». Le terme "solution" donne l'impression de quelque chose de bien défini et de correct. L'ensemble des solutions pour une équation mathématique est certainement bien défini, mais pouvons-nous dire la même chose pour les solutions d'un jeu ? Bien que la solution d'un jeu soit souvent vue comme analogue à la solution d'un problème de maximisation, en fait un concept théorique de solution n'est rien de plus que le traitement d'une interprétation particulière que nous avons choisi de privilégier.

Un autre terme souvent utilisé dans les discussions est celui de "recommandation". Il est souvent dit qu'un concept de solution produit un profil de recommandations qui ne sont pas contre-productives (2). Par exemple, un équilibre de Nash est dit être un profil de recommandations, une pour chaque joueur, avec la propriété que si les joueurs croient que les recommandations sont suivies, aucun n'a intérêt à ne pas suivre ces recommandations.

(2) [Self defeating.]

Je pense que le terme "recommandation" est ici trompeur. Le terme implique "autorité", "instruction" et "la bonne chose à faire". Le terme d'équilibre de Nash peut-il être considéré comme synonyme de recommandation ? Une recommandation doit être formulée par quelqu'un. Qui fait la recommandation dans ce cas ? Est-ce seulement un théoricien des jeux ? Quels sont ses objectifs ? Est-il possible de considérer un équilibre de Nash inférieur au sens de Pareto comme une recommandation ? Et pourquoi faudrait-il que l'auteur des recommandations se limite lui-même aux recommandations qui ne sont pas contraires aux buts recherchés par les joueurs ? Ne serait-il pas préférable, dans un jeu de mille-pattes, de recommander à chaque joueur de ne jamais stopper le jeu, même si ce n'est pas une recommandation contre-productive ?

5. Commentaires finaux

Le lecteur peut avoir l'impression que l'objectif de ce chapitre est de diminuer l'importance de la théorie des jeux et d'en décourager les utilisateurs potentiels. Je pense au contraire que la théorie des jeux se dirige vers de nouveaux champs d'intérêt et sera bientôt le témoin de développements passionnants. Cependant, je ne m'attends pas à ce que la théorie des jeux devienne "pratique" au sens où beaucoup de gens l'entendent. Comme je n'ai jamais compris ces espoirs mis dans la théorie des jeux, je ne suis ni désappointé ni pessimiste.

Je vois la théorie économique analytique comme la recherche de relations entre les concepts, hypothèses et affirmations que nous utilisons pour comprendre les interactions humaines. Une situation stratégique peut être analysée par de nombreuses voies. Typiquement, il y a différents arguments qui permettent de soutenir une suite d'actions ou une autre. Le mieux que nous puissions faire analytiquement (contrairement à l'observation empirique) est de clarifier les relations entre différents types d'analyse. "Etablir des liens", et "comprendre", cependant, n'est pas donner un conseil ou être pratique. A mon sens, parvenir à une meilleure compréhension des arguments utilisés dans les interactions humaines et les raisonnements est, en soi, un projet passionnant et qui en vaut la peine. Pour d'autres points de vue sur la signification de la théorie des jeux, voir Aumann (1987) et Binmore (1990).

Lors d'une conférence où je présentais ce chapitre, un auditeur m'a demandé si je ne me sentais pas frustré par le fait que mon travail et mes points de vue n'apportent pas de contribution au monde. Sans doute le suis-je. Cependant, je ne pense pas que le présent essai soit inutile au sens « d'influencer le monde ». Il y a peu de tâches dans le monde académique des sciences sociales qui soient plus importantes que de combattre l'autorité injustifiée. Ce chapitre est ma modeste contribution à cette importante bataille.

Remarques finales

A l'ouverture de cette courte série de conférences, j'ai présenté brièvement cinq sujets sous l'intitulé *Economie et langage*. Vous avez maintenant réalisé que les sujets étaient en fait assez distincts. J'ai trouvé difficile d'indiquer un point commun entre ces sujets, mis à part le titre et le conférencier.

Je suis profondément reconnaissant aux trois personnes qui ont accepté de commenter ce manuscrit [...].

Les discutants (3) m'ont persuadé que le dernier chapitre [*i.e. ce chapitre 5 NdT*] est de nature très différente des autres. S'il est dans ce livre, au-delà de mon usage rhétorique de l'expression "économie et langage", c'est seulement parce qu'il présente mon approche générale de la théorie économique, qui est un fil directeur de ces conférences. Dans cette approche, la théorie économique :

(3) Il s'agit de Tilman Börgers, Bart Lipman, Johan van Bethem, dont les commentaires figurent en annexe de l'ouvrage (NdT).

- parle plus des méthodes et des questions de modélisation ;
- discute moins de la substance économique ;
- insiste sur la question de la complexité comme élément majeur dans la discussion ;
- se considère comme l'étude des arguments utilisés par les êtres humains.

Faire la lumière sur les arguments utilisés par les hommes dans les interactions sociales et rechercher "la logique de la situation", expression que Karl Popper utilise pour caractériser l'ambition des sciences sociales, est en effet ce que je perçois comme l'objectif de la théorie économique.

Traduit par Michel Hollard.

Références bibliographiques

Aumann R. (1987), « What is game theory trying to accomplish ? », *in* K.J. Arrow et S. Honkapohja (eds), *Frontiers of Economics,* Oxford, Blackwell

Binmore K. (1990), « Aims and scope of game theory », *in* K. Binmore, *Essays on Foundations of Game Theory,* Oxford, Blackwell.

Dixit A.K. et B.J. Nalebuff (1991), *Thinking Strategically,* New York, Norton.

Harsanyi J. (1973), « Games with randomly distributed payoffs: a new rationale for mixed-strategy equilibrium points », *International Journal of Game Theory,* 2, 486-502

Henderson W., T. Dudley-Evans et R. Backhouse (eds.) (1993), *Economics and Language,* London, Routledge.

Mc Closkey D.N. (1998), *The Rethoric of Economics (Rhetoric of Human Sciences),* 2nd ed., Madison, University of Wisconsin.

Mc Millan J. (1992), *Games, Strategies and Managers,* Oxford, Oxford University Press.

Nash J. (1950), « The bargaining problem », *Econometrica,* 18, 155-162.

Radner R. et Rosenthal (1982), « Private information and pure strategy equilibrium », *Mathematics of Operation Research,* 7, 401-9.

Rubinstein A. (1991), « Comments on the interpretation of game theory », *Econometrica,* 59, 909-24.

Rubinstein A., Z. Safra et W. Thomson (1992), « On the interpretation of the nash bargaining solution », *Econometrica,* 60, 1171-86.