Anticiper et choisir

Sylvain BÉAL

Comment expliquer de manière rigoureuse et systématique les décisions prises dans des contextes aussi différents que ceux de la crise des missiles de Cuba ou de la tragédie *Macbeth* de Shakespeare ? Steven Brams recourt à la théorie des jeux pour formaliser et analyser ce type de situations en insistant sur l'importance de mesurer les conséquences à long terme que peuvent engendrer des choix.

Recensé : Steven Brams, *Game Theory and the Humanities: Bridging Two Worlds.* Cambridge, MIT Press, 2011, 336 p.

Steven Brams est professeur de Sciences Politiques à l'Université de New York et publie depuis les années 1960 des travaux en théorie des jeux (définie ci-après), mais également en théorie du vote, et plus généralement encore sur les problèmes d'allocation de ressources rares. S'inscrivant dans la continuité de ses précédents travaux de recherche, son dernier ouvrage *Game Theory and the Humanities: Bridging Two Worlds*, exploite la théorie des jeux pour proposer un éclairage original sur des questions historiques, littéraires, théologique et philosophique, alors que les applications les plus connues et les plus répandues ont plutôt pour cadre l'économie, la sociologie, la science politique et la biologie.

La théorie des jeux

La théorie des jeux est un formalisme né sous l'impulsion de John von Neumann et Oskar Morgenstern dans la première moitié du XXe siècle qui propose de décrire des situations conflictuelles entre des agents (les joueurs). Un jeu est décrit par quatre éléments :

- Un ensemble de joueurs impliqués dans le jeu ;
- Un ensemble de règles du jeu précisant la procédure à suivre et les possibilités d'action offertes aux joueurs ;
- Un ensemble d'issues. Chaque issue est la conséquence des actions prises par les joueurs ;
- Un ensemble de gains. Chaque issue détermine des gains distribués aux joueurs, de sorte à refléter leur niveau de satisfaction.

Dans l'approche non-coopérative, le pouvoir de décision est complètement transféré aux joueurs. Autrement dit, ce sont les joueurs qui décident individuellement et égoïstement. L'étude est centrée sur les comportements décentralisés des joueurs et sur les mécanismes stratégiques et informationnels qui facilitent la coopération.

Un premier exemple de jeu non-coopératif est le jeu d'échecs. Ici, les joueurs sont les deux participants au jeu d'échecs. Les règles du jeu sont celles du jeu d'échecs. Les issues peuvent être échec et mat, le match nul, etc. Les gains peuvent prendre la forme d'un chiffre 1 pour un joueur ayant gagné la partie, -1 dans le cas contraire et 0 en cas de match nul. Un autre exemple naturel de jeu non-coopératif est la concurrence oligopolistique sur les marchés. Ici, les joueurs sont les firmes. Les actions des joueurs sont, par exemple, les prix que peuvent pratiquer les firmes, et les règles du jeu indiquent notamment si les prix sont fixés simultanément ou séquentiellement. Les issues sont les situations de marché spécifiant la part de marché revenant à chaque firme en fonction des prix fixés. Les gains sont les profits ou pertes réalisés par les firmes sur ces issues.

À l'inverse, dans l'approche coopérative, le pouvoir de décision est remis entre les mains de la collectivité. Les joueurs ne décident plus de manière individuelle mais collective. Ils sont supposés coopérer malgré les conflits d'intérêt : la signature d'un contrat de coopération les engage définitivement. L'analyse est centralisée et nettement orientée vers les accords de coopération que les joueurs peuvent signer.

Un exemple de jeu coopératif est la situation d'un aéroport qui va se doter d'une nouvelle piste d'atterrissage que pourront exploiter plusieurs compagnies aériennes impliquées dans le projet. L'aéroport souhaite déterminer la répartition des frais de construction entre ces compagnies aériennes. Ces frais sont proportionnels à la longueur de la piste qui est sur le point d'être construite, et la longueur de la piste dépend de la taille des avions qu'elle accueillera, donc de la flotte de chaque compagnie aérienne. Ici les joueurs sont les compagnies aériennes. Afin de déterminer une juste part du coût de construction qui incombe à chaque compagnie aérienne, les règles du jeu indiquent quel type de piste aurait du être construit si seuls certains sous-ensembles (ou coalitions) de compagnies aériennes s'étaient impliquées dans le projet de construction. Les issues spécifient, pour chaque coalition de compagnies aériennes, quels auraient été les frais de construction de la nouvelle piste si seuls les membres de cette coalition avaient participé à son exploitation. Ces données permettent d'évaluer, dans une certaine mesure, la contribution de chaque compagnie aérienne au coût de construction. Par exemple, associer au projet une compagnie aérienne dotée de petits avions n'augmentera pas le coût de construction de la nouvelle piste, de sorte qu'on peut imaginer qu'une telle compagnie paiera une part modérée du coût total de construction. Les gains sont l'ensemble de toutes les répartitions du coût total de construction entre les compagnies aériennes.

L'ouvrage a pour objectif d'expliquer des faits historiques ou de fictions à travers le spectre de la théorie des jeux. Steven Brams montre que sa *théorie des mouvements* est parfois plus adéquate dans ce contexte que les outils standards de la théorie des jeux comme *l'équilibre de Nash*. L'enjeu est donc double : d'une part, justifier de manière rigoureuse des prises de décisions historiques et fictionnelles, et, d'autre part, mettre en lumière certaines limites de la théorie des jeux standard. Au delà d'un chapitre introductif anecdotique, les neuf autres chapitres possèdent une grande cohérence dans le sens où Steven Brams étudie le plus souvent les questions posées dans chaque chapitre en les modélisant par des jeux non-coopératif à deux joueurs et à deux stratégies par joueurs. Les chapitres 4 à 6 méritent toutefois un traitement séparé dans la mesure où ils font exception à cette règle sur différents aspects.

L'équilibre de Nash – célébré à l'occasion de l'attribution du prix Nobel d'économie en 1994 – est une solution incontournable dans les travaux de recherche en théorie des jeux. Une issue d'un jeu non-coopératif est un équilibre de Nash si aucun joueur ne peut augmenter son gain en changeant de stratégie de manière unilatérale. L'utilisation de ce concept est désormais couramment répandue, en particulier en économie où il est enseigné dès la seconde année de licence. On ne peut probablement pas en dire autant de la théorie des mouvements, dont l'exploitation a été jusqu'à aujourd'hui plus confidentielle. Dans la mesure où les principales conclusions de l'ouvrage s'adossent à cette théorie, il convient d'en présenter les principaux contours.

La théorie des mouvements

Les décisions prises par les joueurs sur un équilibre de Nash sont par nature statiques. Chaque joueur y sélectionne une stratégie optimale compte tenu de la stratégie optimale de ses opposants. Cette propriété de stabilité repose toutefois sur le fait que chaque joueur n'évalue pas les conséquences qu'une modification de sa stratégie pourrait avoir sur ses concurrents. La théorie des mouvements critique ce caractère statique des décisions prises par les joueurs sur un équilibre de Nash. Plus précisément, on peut concevoir que si un joueur remanie sa stratégie, alors cela peut conduire l'un de ses opposants à réviser sa propre stratégie, révision qui pourra à son tour inciter un troisième joueur (potentiellement identique au premier) à altérer sa stratégie et ainsi de suite. C'est l'un des principes sur lequel s'appuie Steven Brams pour construire la théorie des mouvements.

Les règles de la théorie des mouvements

Le point de départ de la théorie du mouvement est un jeu non-coopératif à deux joueurs et deux stratégies, typiquement représenté par une matrice de gains (un exemple est construit dans le tableau 2). La théorie des mouvements est décrite par les règles suivantes (p. 58-61) :

1) Une paire de stratégies (i.e. une case de la matrice) est désignée comme état initial.
2) Chaque joueur a la possibilité de changer unilatéralement de stratégie, conduisant à un nouvel état (i.e. une autre case de la matrice)
3) À la suite d'un changement de type 2), le joueur qui n'a pas agi a la possibilité de changer lui aussi unilatéralement de stratégie, conduisant à un autre état.
4) Les « mouvements » décrits dans les points 2) et 3) se poursuivent jusqu'à ce qu'aucun des deux joueurs ne décide de modifier sa stratégie, auquel cas l'état courant est qualifié d'état final.
5) Un joueur ne change pas de stratégie lorsqu'il préfère l'état courant à l'état final auquel son changement de stratégie pourrait conduire, ou lorsque son changement de stratégie ramène ultérieurement à l'état courant.
6) Chaque joueur est doté d'une rationalité visionnaire dans le sens où ses décisions stratégiques prennent en compte les mouvements que l'autre joueur peut entreprendre en réponse à son propre mouvement, mais aussi les mouvements qu'il peut lui-même effectuer en réponse aux mouvements que l'autre joueur peut entreprendre en réponse à son mouvement initial, et ainsi de suite.

La théorie des mouvements incorpore clairement des aspects dynamiques au processus de prise de décision, en particulier l'idée selon laquelle la rationalité des joueurs n'est pas myope. Les principes inhérents à une telle rationalité ne sont pas nouveaux, et Steven Brams n'est pas le seul à en exploiter la richesse. On en trouve par exemple la trace chez le prix Nobel d'économie John Harsanyi (1974), lorsque l'auteur envisage un processus dynamique de négociations basé sur un système d'offres et de contre-offres. Harsanyi insiste sur le fait qu'un joueur, ou qu'une coalition de joueurs, peut formuler une offre induisant un état intermédiaire moins favorable que le statu quo, sous la contrainte que les réactions des autres joueurs à ce sacrifice conduisent à un meilleur résultat in fine. Le travail d'Harsanyi a notamment inspiré à Joseph Greenberg (1990) sa *théorie des situations sociales*, qui présente plusieurs similitudes avec la théorie des mouvements de Steven Brams et permet d'unifier dans un même cadre d'analyse les jeux non-coopératifs, les jeux coopératifs et les modèles de choix social. La théorie des situations sociales est plus ambitieuse que la théorie des mouvements. Elle permet de traiter des jeux avec un grand nombre de joueurs, avec un grand nombre de stratégies, et admet aussi la formation de coalitions de joueurs. Le cadre formel et les résultats proposés par Greenberg ont ouvert une nouvelle littérature qui s'enrichit considérablement depuis une quinzaine d'années. La théorie des mouvements n'a pas connu à ce jour de développements similaires.

Religion, fiction et faits historiques à l'épreuve de la théorie des jeux

La théorie des mouvements est abondamment utilisée au cours de l'ouvrage pour illustrer comment les décisions prises par des personnages de fiction et des personnages historiques obéissent à certains critères de rationalité. Ces chapitres peuvent être regroupés en trois catégories. Dans la première catégorie, Steven Brams considère Dieu comme un joueur à part entière afin de mettre en lumière certains passages de la Bible ou certains arguments théologiques. À titre d'exemple, l'auteur modélise le pari de Pascal sur la croyance en Dieu. Dans la seconde catégorie, Steven Brams analyse les décisions prises par des personnages de fiction. Plusieurs scènes clés de romans ou de pièces célèbres sont abordées, comme *Macbeth* et *Hamlet*. Dans la troisième catégorie, l'auteur examine des évènements historiques, comme l'interaction entre le président américain Jimmy Carter et l'Ayatollah Khomeiny durant la crise des otages en Iran ou la crise des missiles cubains entre les États-Unis et l'Union Soviétique.

Illustration avec Samson et Dalila

Les Philistins, ennemis d'Israël, soudoient Dalila afin qu'elle découvre le secret de la force de Samson, l'Israélite. Après avoir séduit Samson, Dalila essaie de lui soutirer ce secret. Samson lui répond par un mensonge. Pensant avoir appris son secret, Dalila trahit Samson une première fois. À deux autres reprises, Samson lui ment. Lorsque Dalila lui demande pour la quatrième fois de lui dévoiler son secret, Samson lui apprend que sa force vient de sa chevelure de nazir. Dalila le trahit alors une fois de plus et, durant son sommeil, fait couper ses sept tresses de Samson par sa servante. Elle appelle ensuite des Philistins qui crèvent les yeux de Samson, privé de sa force et de l'aide de Dieu.

Steven Brams modélise l'interaction entre Dalila et Samson (dans les sections 2.4 et 2.5 aux pages 50 et suivantes) par un jeu non-coopératif à deux joueurs et deux stratégies que l'on représente par la matrice des gains suivante :

		Samson	
		Garder son secret	Dévoiler son secret
Dalila	Ne pas harceler	(2 ; 4)	(4 ; 2)
	Harceler	(1 ; 1)	(3 ; 3)

Cette matrice se lit de la manière suivante. Dalila dispose de deux stratégies : Ne pas harceler et Harceler Samson. Samson possède deux stratégies : Garder son secret et Dévoiler son secret à Dalila. Chaque combinaison de stratégies conduit à une issue, et les gains associés reflètent uniquement les préférences des deux protagonistes sur ces issues. Sur une paire de gain, le premier chiffre est le gain perçu par Dalila, le second chiffre le gain perçu par Samson. Un joueur reçoit 4 sur l'issue qu'il préfère parmi les quatre, et le gain diminue jusqu'à 1 pour l'issue qu'il considère comme la pire. Steven Brams justifie longuement la distribution des gains dans la matrice. À titre d'exemple, la combinaison de stratégies (Ne pas harceler, Garder son secret) est la meilleure issue pour Samson (gain de 4), puisqu'il n'a pas été harcelé par Dalila et n'a pas dévoilé son secret. Cette même combinaison de stratégies est la seconde pire issue du point de vue de Dalila (gain de 2). En effet, Dalila préfère avoir découvert le secret de Samson, et, quel que soit le comportement de Samson, elle préfère également ne pas l'avoir harcelé.

L'analyse de ce jeu est double. D'une part, il est facile de repérer que la combinaison de stratégies (Ne pas harceler, Garder son secret) est l'unique équilibre de Nash du jeu. En effet, partant de (Ne pas harceler, Garder son secret), aucun joueur ne peut accroître son gain en choisissant son autre stratégie, étant donné la stratégie de l'autre joueur. Au contraire, partant de toute autre combinaison de stratégies, au moins l'un des joueurs peut améliorer son gain en changeant de stratégie. Par exemple, partant de (Harceler, Dévoiler son secret), il est dans l'intérêt de Dalila de remplacer sa stratégie Harceler par Ne pas harceler : le résultat sera le même (elle découvrira le secret de Samson), mais n'aura pas eu à le harceler pour y parvenir.

D'autre part, Steven Brams mobilise la théorie des mouvements afin de sélectionner comme seule issue plausible la combinaison (Harceler, Dévoiler son secret). Sans trop entrer dans les détails techniques, on peut donner l'intuition de ce résultat. Pour ce faire, partons de l'unique équilibre de Nash (Ne pas harceler, Garder son secret). Dalila peut-elle améliorer son sort en changeant de stratégie ? Dans l'immédiat, la réponse est non puisqu'un tel changement conduit à la pire issue (Harceler, Garder son secret) du point de vue des deux joueurs. Néanmoins, ce changement va inciter Samson à réviser à son tour sa stratégie pour atteindre l'issue (Harceler, Dévoiler son secret). En réponse, Dalila va une fois encore changer de stratégie pour accéder à l'issue (Ne pas harceler, Dévoiler son secret). À ce stade, Samson doit-il revenir à l'issue initiale (Ne pas harceler, Garder son secret) par une nouvelle modification de stratégie ? La réponse est oui. En effet, en procédant de la sorte, un cycle complet aura été effectué. La règle 5 de la théorie des mouvements énonce alors que l'issue (Ne pas harceler, Garder son secret) sera sélectionnée, et il s'agit de issue préférée de Samson. Sachant cela, Dalila comprend qu'elle ne doit pas donner l'opportunité à Samson de pouvoir revenir à l'issue initiale. Autrement dit, elle n'a pas intérêt à générer le mouvement de (Harceler, Dévoiler son secret) à (Ne pas harceler, Dévoiler son secret) car il conduirait à l'issue initiale sur laquelle le gain de Dalila est plus faible que sur l'issue (Harceler, Dévoiler son secret). Ici, l'argument invoqué s'appuie son le concept *d'induction à rebours* souvent exploité en théorie des jeux : un joueur prend une décision sur la base d'un choix optimal que son opposant pourrait être amené à faire par la suite. Ce concept apparaît implicitement dans la règle 6 de la théorie des mouvements. Par le même raisonnement, on peut conclure que Samson préfèrera induire le mouvement de (Harceler, Garder son secret) à (Harceler, Dévoiler son secret), et Dalila préfèrera produire le mouvement initial de (Ne pas harceler, Garder son secret) à (Harceler, Garder son secret). En fin de compte, l'issue (Harceler, Dévoiler son secret) décrite dans la Bible est sélectionnée par la théorie des mouvements.

Les chapitres 4, 5 et 6, dans lesquels la théorie des mouvements n'est pas mobilisée, sont paradoxalement les plus intéressants. Chacun de ces chapitres aborde un problème de choix social sous un angle distinct. Le chapitre 4 traite du problème de répartition équitable d'objets indivisibles. L'analyse théorique nécessite ici moins d'hypothèses que les chapitres précédemment évoqués, tout en autorisant plus de liberté, par exemple sur le nombre des individus participant au problème. Steven Brams démontre par l'exemple l'incompatibilité entre plusieurs propriétés désirables pour la répartition des objets (phénomène qu'il qualifie de paradoxe).

Paradoxe dans le partage équitable d'objets
Pour un problème de juste répartition d'objets indivisibles, on dit qu'une allocation des objets est *efficiente* (ou optimale au sens de Pareto) s'il n'existe pas d'autre allocation plus avantageuse pour certains individus et au moins aussi avantageuse pour les autres. On dit également qu'une allocation des objets *résiste à la convoitise* si chaque individu préfère son panier d'objets au panier d'objets reçu par chacun des autres individus. Steven Brams montre que ces deux propriétés peuvent s'avérer inconciliables (exemple 4.1 page 97). Pour cela, considérons trois individus notés A, B, et C devant se répartir six objets indivisibles notés 1 à 6. Les préférences des individus sont données par :

- Individu A : 1>2>3>4>5>6

- Individu B : 4>3>2>1>5>6
- Individu C : 5>1>2>6>3>4

où la notation « $n>m$ » signifie que l'objet n est préféré à l'objet m.

Tout d'abord, considérons l'allocation attribuant les objets 1 et 3 pour l'individu A, 2 et 4 pour l'individu B et 5 et 6 pour l'individu C. Cette allocation résiste à la convoitise. En effet, pour l'individu A, on a 1>2 et 3>4, de sorte qu'il ne convoite pas le panier de l'individu B, et 1>5 et 3>6, de sorte qu'il ne convoite pas le panier de l'individu C. Il est en de même pour les deux autres individus. Il est également possible de montrer qu'il n'existe pas d'autre allocation des objets résistant à la convoitise. Maintenant, considérons l'allocation alternative qui assigne les objets 1 et 2 pour l'individu A, 3 et 4 pour l'individu B et 5 et 6 pour l'individu C. Le seul changement relativement à l'allocation initiale est les individus A et B s'échangent les objets 3 et 2. Puisque 2>3 pour l'individu A et 3>2 pour l'individu B, on comprend que l'allocation initiale qui résiste à la convoitise ne peut pas être efficiente.

Le chapitre 4 n'illustre aucun fait historique ou de fiction, contrairement à la plupart des autres chapitres de l'ouvrage. Steven Brams aurait probablement pu combler ce manque, en renvoyant par exemple le lecteur au problème d'arbitrage issu du Talmud et étudié, entre autres, par Barry O'neill (1982) et Robert Aumann, prix Nobel d'économie, et Michael Maschler (1985). Dans ces articles, le problème posé est celui de la juste répartition de l'héritage d'un homme criblé de dettes entre ses créanciers. Le chapitre 5 propose de surpasser des dilemmes sociaux en proposant aux joueurs impliqués de voter pour l'alternative qu'ils préfèrent. Le cadre formel est comparable à celui utilisé dans plusieurs autres chapitres, mais permet une extension bienvenue à un nombre arbitrairement grand de joueurs. Le chapitre 6 revient sur certaines décisions de la cour suprême américaine et montre comment calibrer le pouvoir de vote des participants en fonction des coalitions qui sont susceptibles de se former. L'idée naturelle est ici d'exclure *a priori* la possibilité de formation d'une coalition de votants aux idéologies orthogonales, comme par exemple une alliance entre l'extrême droite et l'extrême gauche en France.

Conclusions

La lecture de l'ouvrage de Steven Brams apporte des sentiments contrastés. On s'aperçoit rapidement que l'auteur est très bien documenté. L'originalité des applications de la théorie des jeux est également séduisante, mais pèche probablement par un manque de variété dans la méthode. L'auteur peine en effet à s'écarter de sa théorie des mouvements. La simplicité des jeux auxquels s'applique cette théorie rend malgré tout la lecture particulièrement plaisante. En effet, le traitement des différentes situations n'est pas technique et les arguments utilisés s'appuient la plupart du temps sur du bon sens. Cependant, il n'est pas facile de déceler les réelles nouveautés comparativement à deux ouvrages précédents de l'auteur, *Biblical Games: Game Theory and the Hebrew Bible* et *Theory of Moves* publiés en 1980 et 1994 respectivement. Le lecteur qui voudrait approfondir les liens entre littérature et théorie des jeux peut en revanche consulter les travaux de Michael Chwe, et notamment son dernier ouvrage *Jane Austen, Game Theorist* paru en 2013.

Game Theory and the Humanities: Bridging Two Worlds déçoit un peu plus quant aux hypothèses nécessaires à la description des résultats. Certaines hypothèses comportementales sont parfois un peu fantaisistes, et les limites de la modélisation sont souvent trop apparentes. Les hypothèses simplificatrices sont pourtant en général bien acceptées dans le travail de modélisation. C'est notamment le cas dans le cadre d'applications, en particulier à l'économie où les modèles deviennent rapidement trop sophistiqués pour être appréhendés. Dans l'ouvrage de Steven Brams, ces hypothèses simplificatrices sont par moment trop marquées pour convaincre un lecteur non familier avec la modélisation. Cette dernière limite traduit assez bien le positionnement relativement flou de l'ouvrage de Steven Brams. C'est justement là son principal défaut. On peut légitimement se demander à quelle audience il s'adresse. Quelques notions de bases en théorie des jeux sont nécessaires à la bonne compréhension de plusieurs chapitres, ce qui semble contre-indiquer la lecture par les néophytes. Même si les applications à la Bible et à la littérature sont rafraîchissantes, il faut aussi avouer qu'elles présentent des limites, en particulier la restriction à deux joueurs et deux stratégies par joueurs. En ce sens, l'ouvrage ne s'adresse pas non plus vraiment à un lecteur familier de la théorie des jeux. D'autres applications de la théorie des jeux, par exemple à l'économie industrielle, sont plus importantes et plus pertinentes pour un lecteur qui souhaiterait connaître la portée des travaux en théorie des jeux. L'ouvrage suscite malgré tout la curiosité de réexaminer des faits historiques et de fictions sous l'angle de la théorie des jeux. Il n'est toutefois pas certain que cela suffise à atteindre complètement l'objectif affiché par Steven Brams dans le titre de son ouvrage : créer un pont entre le monde de la théorie des jeux et le monde des sciences humaines.

Références :

- Aumann R.J. et Maschler M., « Game-theoretic analysis of a bankruptcy problem from the Talmud », *Journal of Economic Theory* 36 (1985), 195-213.
- Chwe, M.S.-Y., *Jane Austen, Game Theorist*, Princeton University Press, Princeton, 2013.
- Greenberg J., *The Theory of Social Situations: An Alternative Game-Theoretic Approach*, Cambridge University Press, Cambridge, 1990.
- Harsanyi J.C., « An equilibrium-point interpretation of stable sets and a proposed alternative definition », *Management Science* 20 (1974), 1472-1495.
- O'neill B., « A problem of rights arbitration from the Talmud », *Mathematical Social Sciences* 2 (1982), 345-371.

Publié dans laviedesidees.fr, le 3 février 2014