L'Actualité économique

Rationalité limitée et interactions stratégiques dans les jeux expérimentaux

Antoine Terracol and Jonathan Vaksmann

Volume 92, Number 1-2, March–June 2016

Économie expérimentale : comportements individuels, stratégiques et sociaux

URI: https://id.erudit.org/iderudit/1039874ar
DOI: https://doi.org/10.7202/1039874ar

See table of contents

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)
1710-3991 (digital)

Explore this journal

Cite this article

Terracol, A. & Vaksmann, J. (2016). Rationalité limitée et interactions stratégiques dans les jeux expérimentaux. *L'Actualité économique*, *92*(1-2), 113–149. https://doi.org/10.7202/1039874ar

L'Actualité économique, Revue d'analyse économique, vol. 92, nos 1-2, mars-juin 2016

RATIONALITÉ LIMITÉE ET INTERACTIONS STRATÉGIQUES DANS LES JEUX EXPÉRIMENTAUX*

Antoine TERRACOL
LED - Université Paris 8
antoine.terracol@univ-paris8.fr

Jonathan VAKSMANN
GAINS, Université du Maine
Centre d'Économie de la Sorbonne
Université Paris 1 Panthéon-Sorbonne
jonathan.vaksmann@univ-lemans.fr

Introduction

La théorie microéconomique standard suppose la rationalité des agents, c'est-à-dire que ces derniers vont faire des choix dans le but de maximiser leur utilité[1]. Cependant, dans une situation d'interactions stratégiques, ce qui est optimal pour un agent va dépendre des choix faits par les autres agents. Il est dès lors impossible de considérer le choix d'un agent pris isolément et la résolution du problème passe par l'analyse jointe du problème de décision de tous les agents impliqués.

L'analyse des interactions stratégiques a une longue histoire en économie, en probabilités ainsi qu'en sciences politiques. Dès 1713, James Waldegrave discute d'une stratégie mixte dans un jeu de cartes à deux joueurs (Bellhouse, 2007). Plus tard, Augustin Cournot (1838) propose une analyse pionnière de l'interaction de plusieurs firmes en situation d'oligopole. Il considère que chaque firme choisit son niveau de production en considérant le niveau de production des autres comme étant donné et construit une fonction de meilleure réponse aux choix des autres firmes. Il montre ensuite qu'un équilibre stable peut se mettre en place par un processus itératif d'équilibration.

* Nous souhaitons remercier les éditeurs ainsi que deux évaluateurs anonymes qui ont contribué à améliorer le contenu de cet article. Selon l'usage, les éventuelles erreurs qui subsisteraient restent de notre responsabilité exclusive.

1. La vision de l'agent économique comme étant à même d'effectuer les calculs permettant d'arriver à cette maximisation de l'utilité a été mise en cause par Simon (1955), qui propose de fonder la théorie du comportement individuel sur des bases plus réalistes.

Les situations d'interactions stratégiques ont ensuite été modélisées de façon formelle par von Neumann (1928) et von Neumann et Morgenstern (1944) et se sont principalement concentrés sur les jeux à deux joueurs à somme nulle. La publication ultérieure de Nash (1950) de son concept d'équilibre dans les jeux non coopératifs à *n* joueurs a achevé l'ancrage de la théorie des jeux et de l'analyse des interactions stratégiques dans la science économique. L'équilibre de Nash requiert cependant un fort degré de rationalité de la part des agents. En particulier, ces derniers doivent non seulement être rationnels en donnant une meilleure réponse à leurs croyances, mais la connaissance commune de la rationalité doit également être établie entre les agents, c'est-à-dire que chaque joueur doit savoir que les autres sont également rationnels, que ces derniers savent que les autres sont rationnels, *etc.*[2] À l'équilibre, puisque les agents choisissent des meilleures réponses qui sont mutuellement compatibles, les croyances doivent être correctes. Le statut positif ou normatif de la théorie des jeux et de l'équilibre de Nash a très tôt fait l'objet de débats et la question de la validité de la théorie des jeux, en tant que corpus à même de décrire le comportement « réel » des agents, a vite été posée. Les premières tentatives de validation empiriques des hypothèses et prédictions de la théorie des jeux par Flood (1952, 1958), bien que critiquables par leur méthodologie, ont permis de conclure que les hypothèses habituelles de la théorie des jeux étaient vraisemblablement trop fortes pour espérer décrire les comportements humains.

Par la suite, la montée en puissance de la méthode expérimentale en économie a permis de tester avec plus de rigueur les prémisses et prédictions de la théorie des jeux. Si ces dernières semblent dans certains cas concorder avec les observations empiriques, il suffit parfois d'une petite modification des paramètres pour que cet accord disparaisse (Goeree et Holt, 2001). Dès lors, de nombreux travaux ont cherché à développer des modèles associant puissance explicative des comportements observés et hypothèses moins fortes sur les processus cognitifs des agents. Cet article passe en revue un certain nombre de ces approches, dont le point commun est de supposer une rationalité moins contraignante de la part des agents.

Lorsque l'on cherche à expliquer les comportements observés dans des jeux expérimentaux, deux aspects sont à considérer. Le premier concerne les réactions initiales des agents, c'est-à-dire la façon dont ils choisissent leurs actions lorsqu'ils font face à une situation stratégique donnée pour la première fois. Ces modèles se concentrent sur le processus cognitif des agents et en particulier sur la façon dont ils considèrent la sophistication des autres joueurs et les erreurs faites dans le choix des actions. Ces modèles sont passés en revue dans la section 1 de cet article. Les modèles correspondant supposent par exemple la présence de bruit dans les réponses des agents. Ces réponses bruitées peuvent servir de base à un nouveau concept d'équilibre (modèle de *Quantal response equilibrium*) ou devenir une barrière au calcul des croyances d'ordre supérieur (modèle de *Noisy Introspection*). D'autres modèles proposent de conserver la rationalité individuelle et de meilleures réponses

2. Mais voir Aumann et Brandenburger (1995) ainsi que Polak (1999) sur la question de la nécessité de la connaissance commune de la rationalité pour l'existence d'un équilibre de Nash.

non bruitées, mais limitent la connaissance commune de la rationalité (*k*-rationalisabilité) ou encore proposent que les agents se basent sur une vision des autres joueurs comme étant moins sophistiqués qu'eux même (*Level-k*, kiérarchie cognitive). Ces modèles ont connu un succès empirique certain. En particulier, le modèle de *Quantal response equilibrium* (QRE) et les modèles de *Level-k* et de hiérarchie cognitive se sont montrés très performants. Malgré sa capacité à reproduire un grand nombre de déviations à l'équilibre de Nash, le modèle QRE a néanmoins été critiqué car les hypothèses sous-jacentes semblent imposer une charge cognitive sur les agents peu compatible avec le désir de proposer un modèle de rationalité limitée. Les modèles *Level-k*, reposant sur des hypothèses pouvant apparaître plus réalistes, semblent avoir les faveurs de la littérature plus récente.

Le second aspect concerne les jeux répétés à observation parfaite et information complète. Dans ces situations, les joueurs sont à même d'observer l'histoire passée des actions des joueurs, et les gains associés à leurs choix. Dès lors, les agents peuvent essayer d'inférer les stratégies des autres joueurs sur la base de cette histoire passée. Ce champ de la littérature se concentre sur ces processus d'apprentissage et de formation des croyances des joueurs lorsqu'ils sont à même d'observer l'histoire passée du jeu. Ces travaux sont exposés dans la section 2. Les premiers travaux sur cette problématique se sont inspirés des recherches en psychologie sur les théories de l'apprentissage. Toutefois, ces approches pionnières, basées sur un simple renforcement des actions les plus performantes dans le passé et déconnectées de toute considération stratégique, sont apparues très limitées pour décrire le comportement d'agents dans des jeux. Les économistes ont alors reformulé des théories d'apprentissage pour y intégrer des notions stratégiques et la prise en compte des croyances des individus dans leur optimisation. Les deux approches d'apprentissage, par renforcement et par les croyances, ont coexisté de manière indépendante jusqu'à ce qu'elles soient réunies au sein d'un modèle unifiant, le modèle *Experience-weighted attraction* (EWA). Ce modèle permet donc de quantifier précisément les poids respectifs du renforcement et des croyances dans le processus d'apprentissage des individus. Cependant, l'ensemble des modèles d'apprentissage partagent une faiblesse commune, celle de considérer des agents purement adaptatifs qui ne tiennent pas compte des conséquences de leurs actions sur le comportement de leurs opposants. En d'autres termes, selon ces approches les joueurs négligent les interactions stratégiques. D'autres approches d'apprentissage sophistiqué ont donc été mises au point plus récemment, selon lesquelles les individus sont en mesure d'adopter des stratégies de manipulation du comportement de leurs opposants. Sans pour autant retomber sur le paradigme de rationalité pure, largement contesté dans de nombreuses situations, l'objet de ces approches demeure d'élargir le spectre des comportements considérés dans les jeux en dépassant le cadre restrictif de la non-sophistication à l'extrême qui se limite à des joueurs purement adaptatifs et passifs face à leur environnement.

1. Modèles de réaction initiale hors équilibre

Une première façon de comprendre les stratégies utilisées par des agents dans un cadre de rationalité limitée est de considérer les décisions prises lorsqu'ils font face à un jeu pour la première fois, ou que le jeu n'a pas d'antécédent clair. Un pan de la littérature cherche à expliquer les déviations initiales de l'équilibre de Nash observées dans les jeux expérimentaux et à relâcher les hypothèses sur le degré de rationalité des agents tout en conservant une certaine précision dans leurs prédictions. Les modèles de cette littérature conservent l'essentiel de la rationalité individuelle, mais utilisent la présence de bruit dans les meilleures réponses des agents, des itérations limitées de connaissance de la rationalité et la croyance en une moindre sophistication des autres joueurs. Leur objectif est de proposer une vision crédible du processus cognitif des agents qui soit à même de reproduire les observations expérimentales.

1.1 *Équilibre plus bruit*

Pour donner du sens aux données observées expérimentalement, dans lesquelles on observe de nombreuses déviations par rapport aux prédictions d'équilibre, le plus simple est souvent de considérer que ces déviations sont des « erreurs » faites par l'agent lorsqu'il choisit son action. Ce dernier connaît l'équilibre de Nash, mais se trompe en choisissant son action. Les erreurs sont coûteuses et leur distribution va donc dépendre des différences de paiement espéré entre la stratégie d'équilibre et celles effectivement jouées. Plus les erreurs sont coûteuses en termes de gain, moins les choix vont dévier de l'équilibre. On modélise typiquement les erreurs comme étant issues d'une distribution logistique dans laquelle la probabilité qu'une stratégie soit choisie dépend du coût de déviation correspondant, ainsi que d'un paramètre de sensibilité aux gains.

Une caractéristique de cette approche est que les coûts de déviation sont évalués en supposant que les autres joueurs ne commettent pas d'erreur et choisissent systématiquement leurs stratégies d'équilibre. Les joueurs sont donc supposés être à même de calculer les équilibres de Nash et supposent que les autres joueurs en font autant. De même que pour les raisonnements d'équilibre, la notion d'équilibre plus bruit suppose des capacités cognitives élevées de la part des sujets. Or, les résultats expérimentaux montrent que, hormis dans les jeux les plus simples, les comportements de joueurs semblent incompatibles avec des raisonnements de point fixe ou d'élimination répétées de stratégies dominées à une profondeur trop grande. De plus, ces modèles recèlent une tension entre le degré de rationalité attribué aux joueurs, capables de calculer l'équilibre de Nash, et le fait qu'ils commettent néanmoins des erreurs lorsqu'ils choisissent effectivement leurs actions. Ensuite, la notion d'équilibre plus bruit peut potentiellement expliquer presque toutes les actions en ajustant le paramètre de sensibilité, ce qui le rend peu informatif sur les processus de décision des agents. Enfin, elle implique que l'action d'équilibre a la plus forte probabilité d'être jouée (puisque le coût de déviation correspondant est nul), ce qui empêche d'expliquer les situations où on observe des déviations systématiques de l'équilibre de Nash (Costa-Gomes *et al*., 2009).

1.2 *Rationalisabilité et élimination des stratégies dominées*

Les raisonnements d'équilibre de Nash supposent d'une part que les individus sont rationnels, que la rationalité est connaissance commune, mais également que leurs croyances sont correctes à l'équilibre. Une façon de réduire la charge cognitive associée à un tel modèle de raisonnement des individus est de ne conserver qu'une partie des hypothèses sous jacentes, à savoir la rationalité et la connaissance itérée de la rationalité, et d'abandonner les anticipations rationnelles et la nécessité d'avoir des croyances exactes à l'équilibre.

Les notions de « rationalisabilité » et de « *k*-rationalisabilité » (Bernheim, 1984; Pearce, 1984) sont les produits de cette approche. Commençons par décrire la *k*-rationalisabilité.

La *k*-rationalisabilité suppose la rationalité individuelle et un nombre fini *k* d'itérations de connaissance de la rationalité des joueurs. Une action est 1-rationalisable si elle est la meilleure réponse à un profil d'actions donné des autres joueurs. Elle est 2-rationalisable si elle constitue une meilleure réponse à un profil d'actions des autres joueurs qui soit 1-rationalisable et ainsi de suite. Dans un jeu à deux joueurs, une action sera *k*-rationalisable si elle survit à *k* itérations d'élimination des stratégies strictement dominées (une stratégie sera éliminée si elle est strictement dominée par une stratégie mixte ou par une stratégie pure).

La rationalisabilité est une version plus exigeante pour le processus cognitif que la *k*-rationalisabilité au sens où elle se base sur la connaissance commune de la rationalité (et non pas uniquement sur un nombre fini d'itérations de connaissance de la rationalité des joueurs). En d'autres termes, une stratégie rationalisable est *k*-rationalisable pour tout *k*.

Dans un jeu où l'espace des stratégies est fini, l'application de la rationalisabilité ou de la *k*-rationalisabilité permet de réduire le nombre de stratégies susceptibles d'être choisies par un joueur rationnel en un nombre fini d'itérations. L'ensemble des stratégies restantes est l'ensemble d'actions dites rationalisables (ou *k*-rationalisables). Cet ensemble contiendra les stratégies supportant le ou les équilibres de Nash, mais ne contient pas nécessairement une action unique et peut même, dans certains jeux, correspondre à l'ensemble des actions possibles du joueur.

Le graphique 1 présente des exemples de jeux à deux joueurs dans lesquels les processus de rationalisation aboutissent à des ensembles rationalisables de tailles différentes. Chaque jeu a un unique équilibre en stratégies pures, souligné dans la bimatrice. Dans le jeu *G*1, l'action *T* n'est pas rationalisable car elle est dominée par la stratégie mixte consistant à jouer *M* avec une probabilité de 0,45 et *B* avec une probabilité de 0,55 (en effet, une telle stratégie mixte aura toujours un gain espéré plus élevé que la stratégie pure *T*, quelle que soit la croyance sur les actions du joueur colonne). Il en résulte que l'action *T* n'est pas rationalisable (ni 1-rationalisable) car elle ne peut être choisie par un individu rationnel. Cependant, dans le jeu G1, il est impossible de continuer à éliminer des stratégies dominées car le joueur colonne n'a pas de stratégie dominée, y compris après l'élimination de *T*.

Toutes les actions du joueur colonne sont donc *k*-rationalisables, et l'ensemble des stratégies rationalisables est donc $\{M, B\}$ pour le joueur ligne et $\{l, m, r\}$ pour le joueur colonne. Dans le jeu *G*2, aucune stratégie n'est strictement dominée, et le critère de (*k*-)rationalisabilité ne permet pas de réduire l'ensemble des stratégies : toutes les actions sont (*k*-)rationalisables. Dans le jeu *G*3, au contraire, l'ensemble des stratégies rationalisables se réduit à l'équilibre de Nash $\{M, r\}$. En effet, on voit que *B* est dominée par *T* et n'est donc pas rationalisable. De même on voit que *l* est dominée par *m* et n'est donc pas rationalisable non plus. L'étape suivante consiste à éliminer *T*, puis *m* pour aboutir à l'équilibre de Nash.

Le concept de (*k*-)rationalisabilité est attrayant par sa simplicité, mais son pouvoir prédictif peut s'avérer faible. De plus, la (*k*-)rationalisabilité ne fournit pas de distribution de probabilités sur les actions qui survivent à l'élimination itérée des stratégies dominées, ce qui la rend moins adaptée à l'analyse économétrique de données expérimentales.

GRAPHIQUE 1

EXEMPLES DE JEUX

***G*1**	***l***	***m***	***r***
T	75,90	27,31	55,43
M	90,40	28,35	31,51
B	63,42	65,86	78,26

***G*2**	***l***	***m***	***r***
T	68,62	65,26	61,42
M	34,54	82,70	66,69
B	84,47	26,88	76,84

***G*3**	***l***	***m***	***r***
T	80,10	15,17	22,10
M	15,24	20,28	25,32
B	20,35	14,38	21,72

1.3 *Équilibre avec bruit,* Quantal Response Equilibrium, Noisy Introspection

Les modèles d'équilibre plus bruit présentés dans la section 1.1 permettent d'expliquer les déviations non systématiques de l'équilibre de Nash observées dans les jeux expérimentaux. Dans ces modèles, la probabilité de choix d'une action hors équilibre est inversement proportionnelle au coût de déviation correspondant. Les joueurs vont donc plus probablement choisir une action donnant un gain espéré

élevé qu'un gain espéré faible, sans pour autant choisir systématiquement l'action optimale. Ces fonctions de « réponse quantale » (*quantal response functions*) sont ainsi des fonctions qui assignent au vecteur des gains espérés des actions disponibles un vecteur de probabilités d'actions qui est monotone dans les gains espérés.

Le modèle de QRE de McKelvey et Palfrey (1995, 1998)[3] supposent que les joueurs ont une certaine croyance sur la distribution des actions des autres joueurs. Étant donné cette croyance, on peut déterminer la meilleure réponse correspondante. Les joueurs donnent ensuite une meilleure réponse bruitée, où la probabilité de chaque action est donnée par une fonction de réponse quantale. L'approche du modèle QRE suppose donc une rationalité limitée de la part des agents qui commettent des erreurs en ne donnant pas systématiquement une meilleure réponse à leur croyances. Cependant, le modèle de QRE ne renonce pas à la notion d'équilibre ni aux anticipations rationnelles. En effet, le modèle QRE est un modèle d'équilibre, et ce dernier est atteint lorsque les croyances de chacun sont cohérentes avec les actions (bruitées) des autres joueurs. En d'autres termes, l'équilibre du modèle QRE est un point fixe dans l'espace des distributions d'actions.

Haile *et al.* (2008) ont néanmoins relevé que, dans son cadre le plus général, le modèle de *quantal response equilibrium* pouvait expliquer n'importe quelle distribution d'actions dans les jeux sous forme normale. Autrement dit, le modèle QRE n'est pas falsifiable dans sa forme générale. Une version de ce modèle impliquant des restrictions testables est proposée par Goeree *et al.* (2005).

D'un point de vue empirique, il est généralement supposé que les actions des joueurs suivent une distribution logistique indexée par un paramètre de précision λ[4]. La probabilité pour que le joueur i joue l'action $a_i \in A_i$ est alors donnée par

$$\Pr(a_i) = \frac{\exp\ \lambda \sum_{a_{-i} \in A_{-i}} \Pr(a_{-i})\,\pi(a_i, a_{-i})}{\sum_{a_j \in A_i} \exp\ \lambda \sum_{a_{-i} \in A_{-i}} \Pr(a_{-i})\,\pi(a_j, a_{-i})}$$

où a_{-i} est le profil de stratégies pour les autres joueurs que i, $\pi(a_i, a_{-i})$ est le gain pour i correspondant au profil de stratégies a_i, a_{-i}, et où , $\Pr(a_{-i})$ est la probabilité que les autres joueurs choisissent le profil, a_{-i} Lorsque λ est nul, les joueurs choisissent leurs actions selon une loi uniforme sur l'ensemble des actions disponibles. Lorsque $\lambda \to \infty$, la probabilité de choisir la meilleure réponse tend vers 1, et l'équilibre QRE tend vers un équilibre de Nash.

Prenons l'exemple du jeu du tableau 1, avec $a > 0$.

---

3. Voir également Chen *et al.* (1997).

4. On parle alors de *Logit quantal response equilibrium*, LQRE. Ce dernier a montré une forte capacité à reproduire les distributions d'actions observées empiriquement (Goeree *et al.*, 2005)

TABLEAU 1

EXEMPLE DE QRE

	L	**R**
U	(3,3)	(0,0)
D	(0,0)	(*a*,*a*)

Les équilibres de Nash en stratégies pures sont dans ce cas (*U*, *L*) et (*D*, *R*). Notons μ la probabilité que le joueur colonne joue *R* et σ la probabilité pour que le joueur ligne joue *U*. L'espérance de gain du joueur ligne s'il joue *U* est donnée par $E_\pi[U] = 3\mu$; et est de $E_\pi[D] = a(1-\mu)$ s'il joue *D*. Le modèle QRE avec probabilité logistique spécifie la probabilité pour que le joueur ligne joue *U* comme

$$\Pr(U) = \sigma = \frac{\exp(\lambda \times 3\mu)}{\exp(\lambda \times 3\mu) + \exp(\lambda \times a(1-\mu))}.$$

De même, pour le joueur colonne, on obtient :

$$\Pr(L) = \mu = \frac{\exp(\lambda \times 3\sigma)}{\exp(\lambda \times 3\sigma) + \exp(\lambda \times a(1-\sigma))}.$$

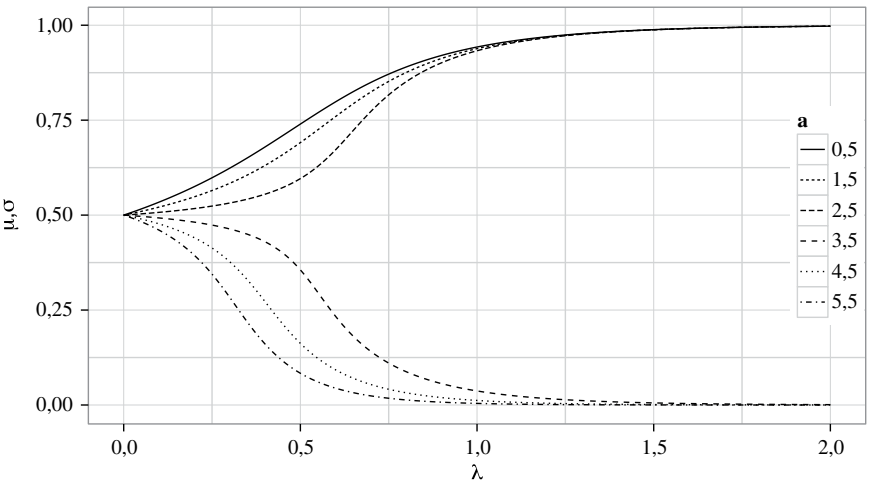
Le jeu étant symmétrique, on a $\mu = \sigma$ et donc

$$\mu = \frac{\exp(\lambda \times 3\mu)}{\exp(\lambda \times 3\mu) + \exp(\lambda \times a(1-\mu))}.$$

En résolvant numériquement cette équation, on obtient les résultats représentés au graphique 2, qui montre l'évolution de la probabilité de jouer *U* (pour ligne) ou *L* (pour colonne) à l'équilibre en fonction de λ, pour diverses valeurs de *a*. On constate que pour des faibles valeurs de λ les joueurs jouent selon une distribution uniforme sur les deux actions, mais lorsque $\lambda \to \infty$, le QRE sélectionne l'équilibre de Nash en stratégie pure, donnant ainsi les gains les plus élevés. La vitesse de cette convergence dépend de la différence entre *a* et 3.

GRAPHIQUE 2

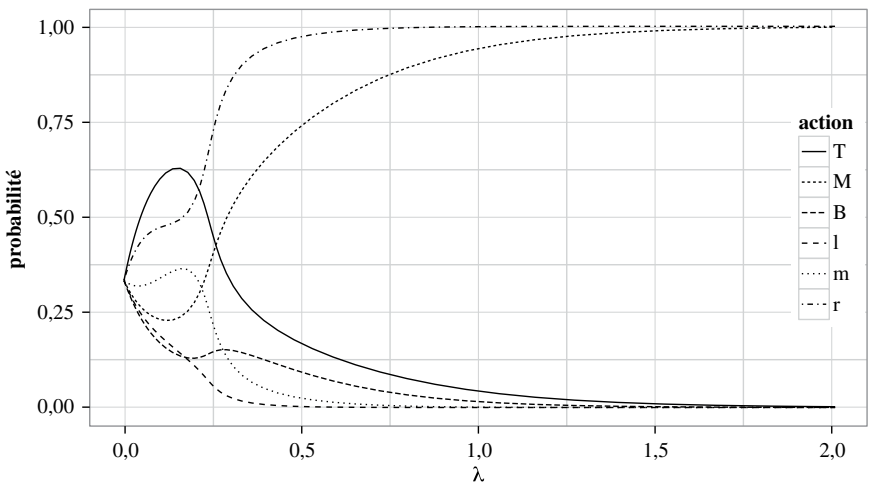
QRE, JEU DU TABLEAU 1



Le graphique 3, quant à lui, présente l'évolution des probabilités du QRE dans le jeu *G*3 du graphique 1. On voit que l'équilibre QRE converge bien vers l'unique équilibre de Nash lorsque $\lambda \to \infty$ grandit. On constate aussi que l'évolution des probabilités de chaque action n'est pas nécessairement monotone.

GRAPHIQUE 3

QRE, JEU G3 DU GRAPHIQUE 1



L'« équilibre avec bruit » du QRE diffère de l'« équilibre plus bruit » de la section 1.1 dans la mesure où le QRE intègre les erreurs des adversaires dans la fonction de réponse de chaque joueur. Au contraire, l'équilibre plus bruit se contente

de bruiter les meilleures réponses des joueurs aux stratégies d'équilibre (non bruitées) de leurs adversaires. D'un point de vue cognitif, les modèles QRE semblent excessivement complexes et coûteux en « calculs » pour les individus. Ces derniers doivent en effet non seulement répondre à une distribution non dégénérée sur les actions de leurs adversaires, mais également trouver un point fixe dans l'espace des distributions d'actions. De ce point de vue, les modèles QRE peuvent être considérés comme plus complexes cognitivement que les modèles classiques d'équilibres sans bruit.

Sur le plan empirique, le modèle QRE a permis d'expliquer un certain nombre de déviations de l'équilibre de Nash; Camerer *et al.* (2004a) affirmant que le QRE devrait remplacer Nash en tant que modèle de référence : « *Quantal response equilibrium, a statistical generalization of Nash, almost always explains the direction of deviations from Nash and should replace Nash as the static benchmark to which other models are routinely compared.* ».

Le modèle QRE a également été utilisé pour expliquer des anomalies empiriques par rapport aux modèles d'équilibres traditionnels dans des jeux sous forme normale tels que les jeux de coordination (Anderson *et al*., 2001), le dilemme du voyageur (Capra, 1999), les enchères (Anderson *et al*., 1998; Goeree *et al.,* 2002), les jeux de participation (Goeree et Holt, 2005), les bulles spéculatives (Moinas et Pouget, 2013) ainsi que dans des jeux de *matching pennies*, de *signaling* et de négociation (Goeree et Holt, 2001).

Goeree et Holt (2004) ont proposé un modèle proche du QRE, mais renonçant à l'hypothèse d'équilibre, tout en conservant le fait que les joueurs répondent à une distribution non dégénérée des actions de leurs adversaires. Ce modèle de *noisy introspection* (NI) construit des croyances d'ordre plus élevé (le premier ordre correspond à la croyance du joueur sur les actions de son adversaire; le second ordre correspond à la croyance du joueur sur la croyance de premier ordre de son adversaire, *etc.*) de façon de plus en plus bruitée. Le modèle NI construit donc la croyance du premier ordre du joueur *i* comme une meilleure réponse bruitée du joueur *j* à la croyance du second ordre de *i*, *etc*. Comme pour le modèle QRE, les erreurs dans le processus de meilleure réponse sont issues d'une fonction logistique dépendant d'un paramètre μ_k gouvernant la précision de la réponse dans la croyance de niveau *k*. Le modèle NI suppose que le bruit μ_k est faiblement croissant au fur et à mesure que l'on construit des croyances d'ordre supérieur. Si $\mu_k = 0$, alors le joueur donne une meilleure réponse non bruitée à ses croyances d'ordre $k + 1$. À la limite, la croyance d'ordre le plus élevé ($\mu_\infty = \infty$) correspond à une distribution uniforme sur l'espace des actions disponibles[5]. Le modèle NI fait des prédictions probabilistes dépendantes de la façon dont le bruit augmente avec l'ordre des croyances. Dans le cas où le paramètre de bruit ne varie pas lors de la construction

5. Cette croyance en une distribution uniforme des actions est en fait atteinte à l'ordre *k* tel que $\mu_k = \infty$. Par exemple, si $_k = \infty \forall k \geq 1$, les individus donnent une meilleure réponse bruitée à une croyance de niveau 1 uniforme.

de croyances d'ordre plus élevé, alors les modèles NI et *Logit quantal response equilibrium* (LQRE) coïncident.

1.4 *Modèles* Level-*k et modèles de hiérarchie cognitive*

Une des critiques apportées aux modèles d'équilibre simple ou aux modèles d'équilibre avec bruit comme le QRE concerne le fort coût cognitif associé à la détermination de l'équilibre. Une autre critique souvent évoquée est que ces modèles ne sont pas construits à partir d'observations empiriques, et sont une construction purement théorique. Les modèles de *Level-k* et de hiérarchie cognitive prennent pour base les observations de Nagel (1995) dans le cadre d'une expérience de concours de beauté[6]. Dans le jeu du concours de beauté, n joueurs doivent simultanément choisir un nombre, habituellement entre 0 et 100. Le gagnant est celui dont le choix se rapproche le plus de p fois la moyenne des choix des n joueurs. Si $p < 1$, le jeu a un unique équilibre de Nash en stratégie pure, qui est de choisir la borne inférieure de l'espace des choix possible, typiquement 0. Cet équilibre de Nash peut s'obtenir par élimination répétée des stratégies dominées. Les données expérimentales obtenues dans de tels jeux montrent que le choix des joueurs se concentrent sur des actions correspondant à des « niveaux de rationalités » croissants[7]. Dans les jeux où l'espace des choix correspond à l'intervalle [0, 100], on observe en effet des pics d'importance décroissante dans la distribution des choix aux niveaux $50p^k$ (33, 22, 15, *etc*. dans le cas où $p = 2/3$. La théorie *Level-k* explique ces pics par la présence de « types » cognitifs dans la population des joueurs. Dans ce modèle, le type de base correspond au niveau 0 (L0), et est composé d'individus non stratégiques choisissant au hasard dans l'espace des choix. Les individus de type immédiatement supérieur, dit de niveau 1 (L1), donnent une meilleure réponse à ceux de niveau 0. Les individus de niveau 2 (L2) donnent une meilleure réponse à ceux de type 1, ceux du type 3 (L3) à ceux du type 2, *etc*.

Considérons le jeu *G*3 du graphique 1. Les joueurs L0 sont supposés choisir chaque action avec une probabilité 1/3. Un joueur ligne L1 va choisir l'action qui maximise son gain espéré face à un joueur L0. Il va donc choisir l'action *T* qui correspond à celle ayant le gain moyen le plus élevé. De même, un joueur colonne L1

---

6. Ce jeu tire son nom de la célèbre citation de Keynes (1936) : *«...professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees. »*

7. Les données de Nagel (1995) sont en fait également compatibles avec un modèle comportant un nombre fini d'éliminations itérées des stratégies dominées, puis de meilleure réponse à une distribution uniforme sur les stratégies restantes. Cependant, des expériences ultérieures (Stahl et Wilson, 1994, 1995) ont permis de trancher en faveur de la théorie *Level-k*.

va choisir l'action *r.* Si on considère maintenant des joueurs L2, ces derniers vont donner une meilleure réponse à des adversaires supposés être L1. Le joueur ligne va donc choisir *M*, qui est une meilleure réponse à *r* et le joueur colonne va quant à lui choisir *m*, qui est une meilleure réponse à *T.* Le cas de joueurs L3 est similaire en ce sens qu'ils vont donner une meilleure réponse à des joueurs L2, si bien que le joueur ligne va choisir *M* et le joueur colonne va choisir *r.* Notons que ces deux joueurs L3 vont choisir une combinaison d'actions correspondant à l'équilibre de Nash. Des joueurs de niveau L4 ou plus élevé vont également choisir la même combinaison d'actions car ils donneront leur meilleure réponse à des actions sous-tendant l'équilibre de Nash.

La théorie *Level-k,* si elle préserve un certain degré de rationalité individuelle et de connaissance itérée de cette rationalité, diffère de la *k*-rationalisabilité en ce qu'elle permet aux joueurs d'être hétérogènes. Les prédictions de la théorie *Level-k*, en termes de distribution des actions, dépendra de la distribution des types parmi les joueurs. Elle se rapproche du modèle NI au sens où les joueurs donnent une meilleure réponse à des joueurs de niveau inférieur[8]. À la différence du modèle NI, le modèle *Level-k* suppose que les joueurs donnent une meilleure réponse à une distribution dégénérée d'actions de leurs adversaires (sauf pour les L1, qui donnent une meilleure réponse à une distribution uniforme, ce qui simplifie grandement le calcul de la meilleure réponse, qui devient celle pour laquelle la moyenne simple des gains est la plus élevée). La théorie *Level-k* propose une représentation du processus cognitif des joueurs vue comme plus réaliste que les modèles d'équilibre tels que le QRE. En effet, elle ne suppose pas de raisonnement de type « point fixe » de la part des joueurs et ne demande qu'un nombre itéré de raisonnement de meilleures réponses (non bruitées) ancrées dans un niveau 0 naïf. Il faut également noter que, dans les applications, il est souvent supposé, à l'instar du modèle d'équilibre plus bruit, que les joueurs commettent des erreurs en choisissant leurs actions et ne donnent donc pas systématiquement une meilleure réponse. Cependant, ces erreurs ne sont pas intégrées que dans l'analyse empirique des jeux (si l'espace des stratégies est continu, il est peu probable que les joueurs choisissent exactement leur meilleure réponse). Le modèle *Level-k* ne suppose pas que les individus prennent les erreurs de leurs adversaires en compte lorsqu'ils forment leurs croyances. Ces dernières sont toujours basées sur une distribution dégénérée des actions du niveau inférieur.

Camerer *et al.* (2004b) proposent un modèle très lié au modèle *Level-k* , le modèle de hiérarchie cognitive (*cognitive hierarchy model*, CH). Ce dernier se démarque du modèle *Level-k* par le fait qu'il ne suppose pas que les individus de niveau k donnent une meilleure réponse à des individus qui sont tous de niveau $k - 1$. Il considère au contraire qu'un individu de niveau $k > 0$ pense faire face à une distribution des niveaux de ses adversaires. Ce dernier considère en particulier que ses adversaires se répartissent sur des niveaux inférieurs au sien. En d'autres termes, le niveau k d'un joueur correspondant au nombre maximum d'itérations

8. L'équivalent du niveau 0 dans le modèle NI correspond à un individu pour lequel $\mu_0 = \infty$. Un individu L1 correspond ainsi au cas $\mu_0 = 0$ et $\mu_1 = \infty$ de la *noisy introspection.*

de meilleure réponses depuis le niveau 0 qu'il peut effectuer et il ne peut concevoir quelle serait la meilleure réponse d'un individu de niveau $k + 1$ à un individu de niveau k (car autrement il serait lui même de niveau $k + 1$). Un joueur de niveau k va donc donner une meilleure réponse à un profil de stratégies de ses adversaires qui va dépendre de sa croyance dans la distribution des niveaux parmi ses adversaires. Par exemple, dans un concours de beauté avec p = 2/3, un individu de niveau $k = 2$ va avoir une croyance (p_0, p_1), $p_0 + p_1 = 1$ sur les proportions de niveaux 0 et 1 parmi ses adversaires. En supposant que les niveaux 0 jouent en moyenne 50, les niveaux 1 vont jouer 33[9,10], et les individus de niveau 2 vont donc donner une meilleure réponse à une action moyenne de leurs adversaires de $50p_0 + 33p_1$. L'introduction d'une hétérogénéité dans la distribution des niveaux des adversaires permet de donner une plus grande flexibilité au modèle car il autorise une distribution non dégénérée des actions d'individus de niveau donné, pour autant que leurs croyances dans la répartition des niveaux ne soient pas identiques. Dans les applications empiriques, il est souvent supposé que la distribution des niveaux suit une loi de Poisson de paramètre λ. La croyance d'un individu de niveau k sera alors donnée par une la distribution de Poisson tronquée à $k - 1$ et renormalisée de façon à ce que la somme des probabilités des niveaux $j < k$ soit bien égale à 1. Notons que, dans le modèle de hiérarchie cognitive, la distribution des niveaux dans la croyance d'un joueur donné n'a pas de raison particulière de correspondre à la distribution réelle des niveau parmi ses adversaires.

De nombreuses études ont mis en lumière le lien entre les capacités cognitives des joueurs et leurs stratégies dans les jeux expérimentaux (Coricelli et Nagel, 2009; Burnham *et al*., 2009; Agranov *et al*., 2012; Brañas Garza *et al.,* 2012) ainsi que le fait qu'ils sous-estiment le niveau de rationalité des autres joueurs (Weizsäcker, 2003), ce qui vient renforcer la crédibilité empirique des modèles *Level-k* et CH. Costa-Gomes *et al.* (2001), Costa-Gomes et Crawford (2006) et Costa-Gomes et Weizsäcker (2008) tentent de distinguer un grand nombre de « types » de joueurs[11] et concluent à la prévalence de types *L*1 et *L*2 dans leurs échantillons. Costa-Gomes et Weizsäcker (2008) trouvent de surcroît que les croyances et les actions des joueurs tendent à ne pas être compatibles.

Les modèles *Level-k* et CH s'ancrent dans le comportement des individus de niveau 0. Ces derniers sont considérés comme des joueurs non stratégiques, c'est-à-dire dont les actions ne sont pas issues d'un comportement de meilleure réponse. Les théories *Level-k* et CH ne précisent pas si ces joueurs existent autrement que « dans l'esprit des joueurs de niveau plus élevé » (Crawford et Iriberri, 2007).

La question de l'existence réelle de joueurs non stratégique de niveau 0 est donc essentiellement empirique. Certain travaux qui cherchent à classifier les joueurs

9. Dans le modèle CH, les individus de niveau 1 considèrent que tous leurs adversaires sont de niveau 0.

10. Pour une « grande » population d'adversaires.

11. Pas uniquement dans le cadre *Level-k*.

par type en étudiant leurs choix de stratégies dans des séries de jeux (Costa-Gomes *et al*., 2001; Costa-Gomes et Crawford, 2006) concluent que la proportion d'individus de niveau 0 est généralement faible ou nulle (voir également Gomes et Crawford, 2006; Crawford *et al.,* 2013). D'autres tels qu'Ivanov *et al.* (2010), Agranov *et al.* (2013) ou encore Burchardi et Penczynski (2014) trouvent une proportion bien plus importante de sujets non stratégiques (de l'ordre du tiers de leur échantillon). Les modèles Poisson-CH estimés par Camerer *et al.* (2004b) impliquent une proportion de joueurs de niveau 0 de l'ordre de 20 % et un niveau moyen de joueurs compris généralement entre 1,5 et 2. Il n'est cependant pas clair si la proportion estimée de niveau 0 dans le modèle Poisson-CH est réelle ou issue des contraintes paramétriques liées à l'utilisation de la loi de Poisson.

Les applications empiriques des modèles *Level-k* et CH ont été nombreuses. Ces modèles ont permis d'expliquer un certain nombre de régularités empiriques de déviation à l'équilibre. On peut en particulier citer Crawford et Iriberri (2007) et la « malédiction du vainqueur » dans les enchères, Costa-Gomes *et al.* (2009) pour les jeux de coordination, Ostling *et al*. (2011) pour les choix de joueurs dans une loterie suédoise et Arad et Rubinstein (2012) pour le jeu dit du « 11-20 ».

1.5 *Remarques conclusives*

Les exigences en termes de processus cognitif des calculs sous-tendant l'équilibre de Nash semblent élevées, et de nombreuses déviations de l'équilibre ont été observées dans les jeux expérimentaux. Un courant de recherche s'est donc concentré sur la possibilité de spécifier des modèles de réaction initiale qui reposent sur des hypothèses moins exigeantes quant à la sophistication des joueurs, tout en permettant de reproduire les régularités observées dans les jeux expérimentaux. Le point commun de la plupart de ces modèles est que, à l'instar du modèle de Nash, ils considèrent les décisions individuelles comme rationnelles, c'est-à-dire comme constituant une meilleure réponse à la façon dont les agents envisagent les actions des autres joueurs. À la différence du modèle de Nash, la façon dont ils envisagent les actions des autres joueurs repose sur un processus imparfait. Une partie de ces modèles considère que les autres joueurs donnent des meilleures réponses bruitées (QRE, NI). Une autre partie suppose que les joueurs posent des limites à la rationalité des autres (*Level-k*, CH) ou qu'il est difficile pour les agents de considérer la connaissance itérée de la rationalité des autres joueurs au delà d'une certaine limite (*k*-rationalisabilité). La pertinence empirique de ces différents modèles est encore une question ouverte. Si le modèle de *Quantal response equilibrium* a connu un succès certain pour expliquer les déviations à l'équilibre de Nash dans un grand nombre de situations, les hypothèses qui le sous-tendent semblent très exigeantes en termes de capacité cognitive des agents, ce qui le rend moins attrayant en tant que modèle de rationalité limitée. Plus récemment, les modèles *Level-k* et de hiérarchie cognitive ont montré leur capacité à expliquer un grand nombre de résultats empiriques, tout en relâchant les hypothèses concernant la complexité des processus cognitifs des agents (Costa-Gomes et Crawford, 2006; Costa-Gomes *et al*., 2009; Crawford *et al.,* 2013).

Les modèles présentés dans cette section se préoccupent des réactions initiales des agents, lorsqu'ils ne disposent pas d'informations autres que celles liées à la structure de la situation stratégique à laquelle ils font face. Dans la section suivante, nous présentons les modèles traitant des phénomènes d'apprentissage, c'est-à-dire le processus par lequel les agents utilisent l'histoire passée du jeu afin d'inférer le comportement de leurs adversaires.

2. Apprentissage

À l'origine, la théorie des jeux ne prenait pas position sur le processus d'équilibration qui demeurait dès lors une énigme fondamentale. Plusieurs mécanismes étaient envisagés. Les modèles d'équilibre général postulent que l'équilibration est issue de variations de prix implémentées par un commissaire walrasien (qui personnalise un processus dynamique non spécifié). Un modèle implicite d'équilibration en théorie des jeux suppose que les agents anticipent spontanément les équilibres de leurs jeux. Les modèles biologiques attribuent l'équilibration à des processus de reproduction génétique, de mutation et de sélection naturelle. Dès les prémisses de la formalisation en théorie des jeux, Nash évoquait une interprétation de l'équilibration en termes d'« action de masse » voisine, dans l'esprit, de la sélection naturelle et qui rappelle également les considérations modernes au sujet de l'évolution culturelle.

Aucune de ces perspectives n'est en mesure d'apporter une description complète quant à la dynamique d'équilibration des êtres humains dans les environnements complexes. En particulier, les hommes apprennent plus rapidement que ce que les modèles biologiques ne prédisent. Ce constat a motivé la recherche de nouvelles approches d'apprentissage. Historiquement, le premier pan de la littérature né des investigations sur ce champ de recherche traitait des propriétés de convergence théoriques vers le ou les équilibres, s'ils existent, de processus dynamiques évolutionnaires ou adaptatifs[12]. Cependant, plus récemment, d'autres approches ont émergé et analysent l'adéquation empirique des modèles d'apprentissage sur des données expérimentales. Leur objectif commun est de retracer aussi précisément que possible, pour chaque décision, comment cette dernière a été prise sur la base du comportement passé du sujet et de son expérience. En d'autres termes, ces approches ont pour but de déterminer, parmi l'ensemble des modèles candidats à retranscrire la dynamique du comportement des joueurs, ceux qui y parviennent le mieux. L'étude de l'apprentissage en théorie des jeux définit différents types de procédures. L'apprentissage peut relever : (i) d'une adaptation au monde extérieur, ou alors (ii) d'une acquisition de connaissances, ou bien encore (iii) d'un mélange de ces deux premières approches. Dans le cas (i), l'apprentissage repose sur l'appréhension de régularités ou de convergences des normes de comportements alors que dans le cas (ii) les joueurs utilisent l'information acquise pour se former des croyances auxquelles ils répondent de façon optimale. En se restreignant aux

12. Le lecteur intéressé pourra par exemple consulter Foster et Vohra (1997) ou Fudenberg et Levine (1998).

modèles les plus couramment utilisés pour analyser l'apprentissage en économie expérimentale, nous présentons dans la suite trois familles de modèles qui illustrent respectivement chacune des trois approches mentionnées ci-dessus.

Une première approche, « l'apprentissage par renforcement », empruntée à la psychologie comportementale, repose sur l'hypothèse que les individus tendent à choisir les actions que se sont montrées les plus payantes dans le passé. En d'autres termes, la propension d'un joueur à opter pour une action donnée repose sur son stock de renforcement dans la mesure où les actions sont renforcées par les paiements passés qu'elles ont rapporté. Un postulat discutable de ces approches, et d'ailleurs débattu par la suite, est que selon cette forme d'apprentissage, les joueurs ne se préoccupent que de leurs gains passés sans s'intéresser à l'historique des actions qui a généré ces paiements.

Confrontés notamment à cette limite lorsqu'ils voulurent à leur tour analyser plus précisément la question de la dynamique d'équilibration dans les jeux, les économistes ont cherché à construire alors des approches plus connectées aux préceptes de la théorie des jeux, mettant ainsi au point une deuxième vague de modèles dits, d'« apprentissage par les croyances ». Cette approche est basée sur le principe que les individus retracent l'historique des choix passés des autres, se forment des croyances sur ce que les autres vont faire et ont ensuite tendance à choisir une meilleure réponse, une stratégie qui maximise leurs paiements espérés étant donnés leur système de croyances.

Enfin, une autre approche d'*experience-weighted attraction* (EWA), combine les particularités des deux précédentes familles de modèles d'apprentissage, en dépit d'ailleurs de leur apparente dissemblance, en rassemblant leurs meilleures caractéristiques dans une même approche qui les incorpore comme cas particuliers. Les théories d'apprentissage par renforcement et par les croyances ont en effet été traitées comme fondamentalement différentes depuis les années 50. Elles étaient généralement perçues comme définitivement inconciliables et ce n'est que récemment, à la fin des années 90 et au début des années 2000, que les chercheurs ont commencé à se poser la question d'une perspective unifiée et se sont mis à travailler sur les connexions entre les formes d'apprentissages concurrentes. Il aura donc fallu attendre près d'un demi siècle pour que soit entreprise la mise en lumière de l'héritage commun qui existait entre ces approches. Le fait que cette question n'ait pas soulevée plus tôt l'attention est probablement dû au fait que les modèles d'apprentissage par renforcement ont été utilisés au départ par les psychologues alors que les modèles d'apprentissage par les croyances ont été mis au point par les théoriciens des jeux et de la décision. Le rapprochement de ces deux disciplines à la fin du 20ème siècle favorisant ainsi la découverte des liens étroits qui unissaient les deux familles de modèles. De plus, ces deux approches reposent sur des principes conceptuellement très différents dans la mesure où les modèles d'apprentissage par renforcement ne reflètent pas la manière dont les autres jouent alors que les modèles d'apprentissage par les croyances ne reflètent pas le succès passé (le renforcement) des stratégies employées.

Cependant, le point commun à tous les modèles d'apprentissage est qu'ils considèrent que les joueurs sont extrêmement non-sophistiqués, ne réagissant qu'à leur propre expérience et voient le comportement de leurs opposants comme généré par un processus exogène de telle sorte qu'ils ne réalisent pas qu'ils peuvent l'influencer. De manière équivalente, selon ces approches, les interactions stratégiques ne jouent pas de rôle dans les jeux. Cette hypothèse est évidemment en complète contradiction avec les fondations mêmes de la théorie des jeux et peut paraître étrange pour bon nombre de théoriciens des jeux.

2.1 *Apprentissage par renforcement*

L'apprentissage par renforcement est historiquement le premier chapitre que compte la littérature sur la formalisation de ce type de processus. Cette approche considère que les individus tendent à actualiser leurs choix au regard du succès que ceux-ci leur ont rapporté dans le passé. Les premières pistes de formalisation sont notamment nées des travaux de Bush et Mosteller (1955), poursuivis plus tard par Cross (1973, 1983) ou encore Harley (1981) et Roth et Erev (1995). Dans les plus récentes approches, les probabilités de choisir chacune de ces actions sont dérivées d'un renforcement cumulatif des paiements reçus procurés par ces actions (et parfois également normalisés comme dans Arthur, 1991).

Les premiers modèles d'apprentissage par renforcement ont été mis au point par les tenants de la psychologie comportementale dans les années 1920. Ce pan de la recherche en psychologie était alors émergeant et allait connaître des développements très importants, notamment jusque dans les années 1960. Lorsqu'ils s'intéressaient aux processus cognitifs descriptifs du comportement des individus, les « behavioristes » se heurtaient alors à la vision « mentaliste » qui ne fournissait qu'une interprétation imprécise de ces processus et qui surtout ne reposait pas sur l'observation directe des choix. Réprouvant fortement cette vision, les behavioristes ont ainsi entrepris de bannir ces formulations vagues des systèmes de cognition pour se concentrer sur des approches susceptibles d'être directement observées dans leurs données. C'est ainsi q'ils prirent comme hypothèse que les déterminants du choix des agents étaient basés sur le renforcement de leur expérience passée.

On dénombre plus précisément trois vagues de formalisation de l'apprentissage par renforcement en théorie des jeux et de la décision. La première vague historique, initiée notamment par Bush et Mosteller (1955) modélise des formes de renforcement basées sur des heuristiques simples et appliquées au champ de la décision dans un environnement parcimonieux. De plus amples développements, toujours dans des cadres de décisions individuelles, furent analysés par Cross (1973, 1983) qui déploya entre autres ces approches dans un contexte de décision en économie. Il est à noter que ces travaux innovants n'ont pas déclenché d'émules pendant plusieurs années, jusqu'aux travaux de Mc Allister (1991), Mookherjee et Sopher (1994, 1997), Roth et Erev (1995) et Sarin et Vahid (2001) parmi d'autres.

Venons-en à présent à une description formelle des modèles d'apprentissage par renforcement. Le « stock de renforcement » a un impact sur l'« attraction » associée à une action qui définit ensuite la probabilité avec laquelle l'agent choisit cette action. L'attraction initiale du joueur i pour l'action $a \in S_i$ est notée $A_i^a(0)$. S_i représentant l'espace des actions du joueur i. Cette attraction initiale peut être postulée *a priori* ou bien estimée sur les données. Le renforcement se fait selon deux principes détaillés formellement comme suit :

$$A_i^a(t) = \phi.A_i^a(t-1) + 1_{\left(a_i(t)=a\right)}.\pi_i(a, a_{-i}(t))$$

où $a_i(t)$ représente l'action choisie par le joueur i à la période t et $a_{-i}(t)$ représente le profil d'action des opposants du joueur i à la période t. $\pi_i(a, a_{-i}(t))$ est le paiement du joueur i lorsqu'il choisit a alors que le profil d'actions de ses opposants est $a_{-i}(t)$. $1_{\left(a_i(t)=a\right)}$ est une fonction indicatrice qui prend la valeur 1 quand le joueur i choisit l'action a à la période t et 0 sinon.

D'autres modèles de renforcement font l'hypothèse que les paiements passés sont moyennés plutôt que cumulés (Mc Allister, 1991; Mookherjee et Sopher, 1994, 1997; Sarin et Vahid, 1999). Dans ce cas, le renforcement est actualisé de la manière suivante :

$$A_i^a(t) = \phi.A_i^a(t-1) + (1-\phi).1_{\left(a_i(t)=a\right)}.\pi_i(a, a_{-i}(t)).$$

Tel que mentionné plus haut, les attractions servent à déterminer les probabilités de choix d'actions des joueurs. La probabilité pour que le joueur i choisisse l'action a à la période $t+1$, notée $P_i^a(t+1)$, doit être monotone croissante en $A_i^a(t)$ et décroissante en $A_i^{a'}(t)$ (où a ≠ a′). La forme principale utilisée dans la littérature est la forme dite logistique, donnée par :

$$P_i^a(t+1) = \frac{\exp \lambda A_i^a(t)}{\sum_{z \in S_i} \exp \lambda A_i^z(t)}.$$

Le paramètre λ mesure la sensibilité des joueurs à leurs attractions. Dans cette fonction de probabilité, l'argument de l'exponentielle au numérateur est simplement l'effet pondéré de l'attraction de l'action a, $\lambda A_i^a(t)$, sur la probabilité de choisir cette même action. Cette forme fonctionnelle a été utilisée dans l'étude de l'apprentissage dans les jeux par Mookherjee et Sopher (1994, 1997), Ho et Weigelt (1996) ou encore Fudenberg et Levine (1998).

L'étude de Roth et Erev (1995), certainement la plus citée en économie parmi celles qui mobilisent l'apprentissage par renforcement, déploie cette approche pour tenter d'expliquer les différences observées dans trois jeux archétypaux. Le jeu de l'ultimatum, le jeu de l'ultimatum avec concurrence parmi les proposants (jeu de marché) ainsi qu'une variante du jeu de biens publics (*best-shot public good games*). Dans tous ces jeux, les prédictions à l'équilibre sont extrêmement inégales. Les observations expérimentales en revanche tendent à montrer que les comportements des sujets convergent vers des divisions très inégales du surplus dans les jeux

d'ultimatums avec concurrence des proposants et dans les jeux de *best-shot public good games*, alors qu'ils convergent plus favorablement vers vers un partage à peu près égalitaire dans les jeux d'ultimatum classiques. Roth et Erev cherchaient à reproduire ces régularités avec un modèle d'apprentissage par renforcement. Ils montrent que le renforcement des paiements passés est en mesure d'approximer correctement la direction de l'apprentissage dans les jeux d'ultimatum et de *best-shot* mais appréhendait mal la vitesse d'apprentissage, trop peu réactive dans les modèles de renforcement. De plus, dans le jeu de l'ultimatum avec concurrence entre proposants, l'adéquation du modèle restait globalement très peu satisfaisante. À la lumière de ces résultats, deux pistes de recherche ont été explorées par la suite. Une partie des travaux entrepris plus tard ont tenté de trouver de nouvelles approches plus robustes qui ne présentaient pas les faiblesses empiriques des modèles de renforcement. Une autre consistait à prendre acte du fait que ce modèle puisse avoir parfois des défaillances et à circonscrire alors des environnements dans lesquels le renforcement fournissait une description adéquate des données.

En résumé, les modèles d'apprentissage par renforcement, au sein desquels seules les actions choisies sont effectivement renforcées, prédisent la direction de l'apprentissage d'une manière relativement correcte. Cependant ces modèles se sont avérés trop lent pour rendre compte du rythme selon lequel les êtres humains apprennent. Une limite des approches de renforcement est que, parce qu'elles postulent que les joueurs ne se préoccupent que de l'historique de leurs paiements, elles supposent donc que les individus négligent beaucoup l'information qui peut leur être fournie dans des environnements riches. Elles s'appliquent ainsi plutôt à des environnements où l'information est limitée, fréquemment soumise à des variations brusques ou à des environnements très changeants. En somme, les faiblesses de ces approches pour ce qui est de décrire le processus d'apprentissage des agents peuvent être imputées au fait que, dans des environnements riches en information, les sujets basent leurs choix sur un ensemble d'information plus large que ne le postulent les modèles d'apprentissage par renforcement. Un moyen d'accélérer l'apprentissage dans ces approches est d'adopter des algorithmes plus sophistiqués afin d'approximer de façon plus adéquate l'apprentissage humain en y incorporant une dimension d'exploration. Cette objectif peut être atteint si l'on considère que les joueurs réévaluent leurs attractions pour des actions non choisies au regard des paiements qu'elles auraient pu leur procurer ou encore en renforçant des actions « similaires » à celles qu'ils ont choisi. Les modèles détaillés par la suite étendent la notion d'apprentissage dans cette perspective.

2.2 *Modèles d'apprentissage par les croyances*

Dans un modèle d'apprentissage par les croyances, les joueurs tendent à opter pour des stratégies qui leur octroient un paiement espéré important étant donné leurs conjectures, formées sur la base de l'observation du comportement passé de leurs opposants.

Historiquement, le premier modèle de croyances remonte à Cournot (1838). Dans ce modèle, les joueurs choisissent une meilleure réponse au comportement de la

période passée. Plus d'un siècle plus tard, les théories de *fictitious play* ont été proposées par Brown (1951) et Robinson (1951). Cette approche suppose que les joueurs forment leurs croyances sur la base de la fréquence des choix passés observés chez leurs opposants. À l'origine, le propos de ces modèles était de calculer des équilibres de Nash de manière algorithmique, fournissant ainsi une forme de rationalisation sur le processus cognitif menant à l'équilibration par tâtonnement successifs.

Les premières observations ont montré que, lorsque les croyances *fictitious play* convergent, elles convergent vers un équilibre de Nash. Par la suite, Shapley (1964) a montré que dans un jeu à somme nulle particulier avec trois stratégies, le processus de *fictitious play* faisait des cycles autour des stratégies sans converger vers l'unique équilibre en stratégies mixtes. Ceci nourri les espoirs de voir en la théorie du *fictitious play* un processus de formation de croyances empiriques capable de converger vers l'équilibre de Nash de manière générale. Les recherches sur les dynamiques d'apprentissage connurent ensuite un coup d'arrêt d'une quinzaine d'années.

Ce n'est que plus tard que le *fictitious play* fut dépoussiéré et reinterprété comme une théorie de l'apprentissage sur la base des comportements effectifs observés plutôt qu'un processus de formation de croyances sur la base de simulations mentales (Fudenberg et Kreps, 1993, 1995). L'une des réévaluations notables qu'ont fournies ces approches qui ont permis d'améliorer le pouvoir prédictif du *fictitious play* est l'introduction de lissage dans le comportement des individus par l'introduction de petits tremblements au sein du modèle de *smooth fictitious play*. Par la suite, Brandts et Holt (1996) et Cooper *et al*. (1997) ont utilisé le *fictitious play* dans des jeux de signaux. Boylan et El-Gamal (1993), quant eux, ont comparé les processus de croyances *fictitious play* et Cournot dans des jeux de coordination et des jeux rationalisables; leurs résultats supportent nettement l'approche *fictitious play*.

Lorsque l'accent fut porté de manière plus importante sur les processus de formation de croyances, et que les approches candidates à décrire les croyances commencèrent à abonder, la question se posa de trouver un cadre susceptible d'unifier les recherches et de fournir des modèles descriptifs englobant. Sur ce point, une classe assez large de processus est incorporée dans l'approche de *fictitious play pondéré* chez Cheung et Friedman (1997). Ces auteurs définissent un modèle qui admet comme cas particuliers, à la fois le modèle *fictitious play* classique (où toutes les observations passées sont pondérées équitablement) et le modèle de Cournot. Plus précisément, dans un jeu à n joueurs, la croyance du joueur i concernant la proportion d'actions k qui sera jouée en $t + 1$ est donnée par

$$B_i^k(t+1) = \frac{r^k(t) + \sum_{u=1}^{t-1} \gamma^u r^k(t-u)}{1 + \sum_{u=1}^{t-1} \gamma^u}$$

où $r^k(t) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} 1_{\left(a_j(t)=k\right)}$.

Notons que dans le jeu à deux joueurs que nous analyserons dans la suite, peut se réécrire

$$B_i^k(t+1) = \frac{1_{\left(a_j(t)=k\right)} + \sum_{u=1}^{t-1} \gamma^u 1_{\left(a_j(t-u)=k\right)}}{1 + \sum_{u=1}^{t-1} \gamma^u}, \quad j \neq i.$$

L'action choisie à une période donnée est escomptée au taux $\gamma \in 0,1$[13]. Lorsque $\gamma = 0$, ce modèle se réduit au modèle de Cournot, où la croyance actualisée à la date t sur l'action a vaut 1 si l'action a été choisie à la période t–1 et 0 sinon. À l'inverse, lorsque $\gamma = 1$, ce modèle se réduit à du *fictitious play*, où la croyance concernant une action donnée correspond à la fréquence avec laquelle cette action a été jouée depuis la première période.

Ainsi, l'attraction du joueur i pour l'action a correspond à son paiement espéré :

$$A_i^a(t) = \sum_{z \in \bigcup_{\substack{j=1 \\ j \neq i}}^{n} S_j} B_i^z(t) \pi_i(a,z).$$

Comme précédemment, les attractions sont ensuite introduites dans une fonction logistique pour déterminer les attractions du joueur i dans l'esprit de l'équation.

Cheung et Friedman (1997) estiment leur modèle sur données individuelles issues de quatre jeux (faucon-colombe, chasse au cerf, acheteur-vendeur et bataille des sexes). Ils trouvent une hétérogénéité substantielle entre les sujets mais également une stabilité des estimations paramétriques entre les différents jeux. Le paramètre γ médian est compris entre 0,25 et 0,50 selon les jeux, donc plutôt dans la région des comportements de type Cournot que *fictitious play* (estimations plus proches de 0 que de 1).

Plutôt que d'avoir recours à des instruments empiriques (comme le *fictitious play* pondéré) pour décrire le processus de formation de croyances des joueurs, plusieurs études en économie expérimentale mesurent directement les croyances des sujets en utilisant des règles incitatives pour induire les joueurs à les reporter de manière véridique. Dans une étude pionnière, McKelvey et Page (1990) élicitent les croyances pour tester l'agrégation d'informations. Il convient également de noter que Camerer *et al.* (2002b) reportent une recherche de Camerer et Weigelt (1988) où les auteurs utilisent des croyances élicitées pour tester les raffinements de prédictions concernant les conjectures hors-équilibres dans des jeux de signaux et d'investissement. Camerer et Karjalainen (1994) trouvent que les croyances élicitées dans un jeu de coordination peuvent être superadditives, reflétant une aversion pour l'ambiguïté liée à l'incertitude stratégique. Plus tard, Nyarko et

13. Le modèle de *fictitous play* pondéré est également appelé modèle de croyances γ-pondérées.

Schotter (2002) utilisent les croyances révélées pour répondre à une question évidente et qui restait pourtant en suspens : quel est le modèle de formation de conjectures qui décrit le mieux les croyances véritables des agents ? Pour répondre à cette question, les auteurs utilisent un jeu 2 x 2 avec un unique équilibre de Nash en stratégies mixtes. Les sujets jouent soixante répétitions de ce jeu et quatre différentes conditions qui croisent appariement fixe *versus* aléatoire avec ou sans révélation de croyances sont implantées. Tout comme d'autres études expérimentales conduisant des jeux avec un unique équilibre de Nash en stratégies mixtes, les fréquences effectives d'actions jouées au travers des soixante répétitions se situent entre la parfaite répartition aléatoire et les probabilités induites par l'équilibre en stratégies mixtes.

Parce que les croyances sont mesurées directement, il est possible d'estimer le paramètre de *fictitious play* pondéré qui fournit la meilleure adéquation aux croyances reportées par les sujets. Les croyances déclarées peuvent ainsi être substituées à $B_i^z(t)$ dans l'équation. Les résultats de Nyarko et Schotter (2002) montrent que le *fictitious play* est en réalité une mauvaise approximation des croyances déclarées dans leurs jeux. En effet, les croyances peuvent entre autres être construites sur davantage d'informations que ne le postule le modèle de *fictitious play* et peuvent par voie de conséquences être plus sophistiquées que ne le laisse penser cette approche. Les sujets sont notamment susceptibles d'anticiper le processus d'apprentissage de leurs opposants. Cette dernière critique va notamment être la pierre angulaire de l'argument qui sera développé dans la section 2.4. De plus, l'utilisation de croyances révélées permet de calculer les probabilités de choix des agents sur la base des paiements espérés « réels ». Ce faisant, Nyarko et Schotter (2002) montrent que le modèle d'apprentissage par les croyances, avec croyances révélées, supplante toutes les autres approches usuelles utilisées dans l'éventail des modèles statistiques d'apprentissage. Ce dernier modèle servira donc naturellement de base à la réévaluation des modèles d'apprentissage qui sera proposée dans la suite de ce papier.

Plusieurs études ont tenté de faire la synthèse des approches d'apprentissage par renforcement et par les croyances. Il s'avère notamment que dans les jeux à équilibres en stratégies mixtes, les modèles de renforcement permettent généralement une meilleure adéquation que les modèles d'apprentissage par les croyances (Erev et Roth, 1998; Mookherjee et Sopher, 1994). Cependant la gamme d'environnements au sein duquel la supériorité du renforcement semble avéré reste cirsconscrite à ce type de jeux. En effet, dans le jeu de coordination analysé par Ho et Weigelt (1996) et Battalio *et al*. (2001) par exemple, l'apprentissage par les croyances fournit une meilleure description des comportements. Par ailleurs, certaines études qui estiment les contributions relatives des différents modèles (Erev et Roth, 1998; Battalio *et al*., 2001) trouvent que la composante croyances est à peu près dix fois plus importante que la composante renforcement.

Ceci étant, il n'en reste pas moins délicat de tirer des conclusions définitives quant aux avantages comparatifs de l'une ou l'autre des approches. D'une part, parce que les gammes de jeux testés diffèrent et de manière plus importante parce que les détails de la spécification et de l'implantation des modèles divergent

également. Au gré des recherches menées, certains paramètres des modèles d'apprentissage par renforcement sont par exemple tantôt ajoutés, tantôt supprimés et la manière dont les attractions sont renforcées est elle aussi fluctuante. Les modèles de croyances sont eux aussi sujets à certaines variations d'une étude à l'autre. La définition des croyances initiales ou encore celle des attractions peut être appréciée différemment selon les analyses. De la même façon, ces études retiennent certaines fois le modèle de Cournot comme approximation, d'autres le *fictitious play*, dans sa version de base ou dans sa version pondérée. Par ailleurs, les tests statistiques utilisés peuvent également varier. La solution naturelle à ce manque d'homogénéité est d'utiliser une plus grande variété de jeux et de statistiques et d'avoir recours à des modèles plus généraux englobant le plus grand nombre possible de modèles développés précédemment comme cas spécifiques. Il serait ainsi possible d'évaluer formellement au sein d'une même approche l'étendue de tel ou tel type d'apprentissage dans le comportement des joueurs. C'est précisément la solution retenue par Camerer et Ho (1999) dans l'approche détaillée dans la prochaine section.

2.3 *Modèle hybride d'apprentissage : modèle d'*Experience Weighted Attraction

Le modèle modèle d'*experience-weighted attraction* (EWA) de Camerer et Ho (1999) repose sur deux variables qui sont actualisées à chaque période. La première, notée $N(t)$, s'interprète comme le « stock d'expérience » accumulé par l'agent. La seconde variable importante est $A_i^a(t)$, l'attraction du joueur *i* pour l'action *a* après que la période *t* ait eu lieu.

Les variables $N(t)$ et $A_i^a(t)$ sont initialisées à $N(0)$ et $A_i^a(0)$. Ces valeurs initiales peuvent être vues comme reflétant l'expérience que l'individu a accumulé avant même que le jeu ne commence, par transferts d'apprentissage depuis des situations similaires expérimentées auparavant. Ces valeurs peuvent également être interprétées comme les résultantes de raisonnements introspectifs au moment d'aborder le jeu.

L'actualisation dans le modèle EWA est gouverné par deux règles. La première actualise le niveau d'attraction. Cette règle d'actualisation prend en compte soit le paiement qu'une action donnée a pu rapporter, soit le paiement que cette action aurait pu rapporter au joueur. Ainsi le modèle EWA tient compte, non seulement des paiements factuels, mais également des paiements potentiels. Plus précisément, ce modèle pondère les paiements hypothétiques que les actions non choisies auraient rapporté par un paramètre δ et pondère les paiements effectivement reçus de l'action choisie $a_i(t)$ par 1 – δ. Le nouveau paiement pondéré peut alors s'écrire $\delta+(1-\delta).1_{\left(a_i(t)=a\right)}\ .\pi_i\left(a,a_{-i}(t)\right)$. La règle d'actualisation des attractions donne $A_i^a(t)$ comme la somme de l'attraction précédente $A_i^a(t-1)$ dépréciée, plus le paiement pondéré à la période *t*, normalisée par le stock d'expérience actualisé $N(t)$. Ce qui conduit à la formule suivante :

$$A_i^a(t)=\frac{\phi.N(t-1).A_i^a(t-1)+\ \delta+(1-\delta).1_{\left(a_i(t)=a\right)}\ .\pi_i\left(a,a_{-i}(t)\right)}{N(t)}.$$

Le taux de dépréciation ϕ représente une combinaison d'oubli et de réévaluation de la manière avec laquelle le comportement des autres joueurs est pris en compte, de sorte que les observations passées deviennent obsolètes et doivent être ignorées. Lorsque ϕ est relativement petit, les joueurs déprécient les observations anciennes plus rapidement et réagissent davantage aux observations les plus récentes. Au final, les attractions du modèle EWA sont ici des moyennes pondérées des attractions passées et des paiements factuels ou des paiements hypothétiques.

La deuxième règle actualise le stock d'expérience selon l'équation suivante :

$$N(t) = (1-\kappa)\phi.N(t-1)+1, \ t \geq 1.$$

Le paramètre κ détermine le taux de croissance des attractions qui reflète la vitesse avec laquelle les joueurs convergent vers une stratégie. Lorsque $\kappa = 0$, les attractions sont des moyennes pondérées des attractions et des paiements passés (avec les poids respectifs $\phi.N(t-1)/[\phi.N(t-1)+1]$ et $1/[\phi.N(t-1)+1]$), de telle manière que la valeur des attractions ne peut dépasser les limites des paiements du jeu. Lorsque $\kappa = 1$, les attractions se cumulent et peuvent alors dépasser les paiements du jeu.

Le terme de paiements pondérés $\delta+(1-\delta).1_{\left(a_i(t)=a\right)} .\pi_i\left(a,a_{-i}(t)\right)$ revêt une importance cruciale dans le modèle EWA. Les attractions de « toutes » les actions non choisies sont actualisées par fois le paiement que l'action « aurait » rapporté si elle avait été effectivement choisie. L'action choisie $a_i(t)$ est actualisée par une fraction additionnelle $1 - \delta$ des paiements qu'elle a effectivement rapporté.

Selon les différentes restrictions de paramètres, le modèle EWA se réduit au modèle d'apprentissage par renforcement ou au modèle de *fictitious play* pondéré. Quand $\delta = 0$, $\kappa = 1$ et pour $N(0) = 1$ (ou de manière équivalente pour $N(t) = 1$, les attractions sont actualisées par

$$A_i^a(t) = \phi.A_i^a(t-1)+1_{\left(a_i(t)=a\right)}.\pi_i\left(a,a_{-i}(t)\right)$$

qui est en fait une version du modèle de renforcement cumulatif. Lorsque κ vaut 0 au lieu de 1, les attractions sont des moyennes pondérées par $\phi / (\phi + 1)$ et $1 / (\phi + 1)$, plutôt que des valeurs cumulées.

Lorsque $\delta = 1$ et $\kappa = 0$, la règle d'actualisation devient

$$A_i^a(t) = \frac{\phi.N(t-1).A_i^a(t-1)+\pi_i\left(a,a_{-i}(t)\right)}{\phi.N(t-1)+1}.$$

On peut montrer (Camerer et Ho, 1999[14]) que cette équation est équivalente à celle qui gouverne l'actualisation du *fictitious play* pondéré. Ainsi, le modèle de *fictitious play* pondéré est un cas particulier du modèle EWA dans lequel les

14. Brièvement, l'astuce consiste d'abord à écrire les croyances à la période t comme une fonction des croyances à la période $t - 1$. Lorsque ces croyances sont utilisées pour calculer les paiements espérés et que les paiements espérés à la période t sont écrites comme une fonction des paiements espérés à la période $t - 1$, le terme de croyance disparaît. Ainsi, l'impact algébrique de l'actualisation des croyances est compris dans le renforcement des actions effectives ou hypothétiques, comme dans l'équation (4).

attractions initiales sont basées sur les paiements espérés, les actions sont actualisées de manière équivalente selon qu'elles sont hypothétiques ou effectives et les attractions passées sont pondérées avec les renforcements courants.

La spécification du modèle EWA fait apparaître l'étroite relation qui existe entre deux familles de modèles, historiquement considérées comme fondamentalement disjointes (Selten, 1991). Certains auteurs ont modifié les modèles de renforcement pour y inclure une composante contrefactuelle par les paiements hypothétiques (Mc Allister, 1991) ou les paiements hypothétiques les plus élevés (Roth et Erev, 1995) mais sans relever toutefois que ces élaborations effectuent la jonction entre les modèles d'apprentissage par renforcement et par les croyances.

La non-linéarité entre les trois paramètres du modèle EWA constitue sa force par rapport aux modèles d'apprentissage par renforcement ou par les croyances pris séparément, ou même par rapport à une combinaison de ces deux derniers modèles. En effet, dans le modèle EWA, les apprentissages par renforcement et par les croyances diffèrent selon trois dimensions : le poids des attractions initiales $N(0)$, le poids des paiements hypothétiques dans l'actualisation des attractions (paramètre δ) et le fait que les attractions puissent ou non dépasser les limites des paiements possibles (paramètre κ). Le modèle EWA n'est donc pas une simple combinaison des modèles d'apprentissage par renforcement et par les croyances puisque ces trois dimensions sont contrôlées séparément. Ainsi, l'estimation du modèle EWA est potentiellement supérieure à la conduite de régressions des choix individuels sur une combinaison pondérée des modèles de renforcement et croyances. Par exemple, une combinaison convexe selon laquelle les paiements espérés auraient un poids δ et les renforcements auraient un poids $1 - \delta$ actualiserait les attractions à la manière du modèle EWA mais sans permettre de prendre en compte une large plage de valeurs possibles pour les attractions initiales ainsi que pour les facteurs d'escompte de l'expérience et de croissance qu'incorpore en plus l'approche EWA[15].

Cependant, le paramètre le plus important du modèle EWA reste le paramètre δ qui fait le lien entre les approches de renforcement et croyances. Plus précisément, celui-ci mesure le poids relatif que les individus accordent aux paiements hypothétiques par rapport aux paiements effectifs dans l'actualisation des attractions. En d'autres termes, il capture les deux lois typiques de l'apprentissage que l'on qualifie de « loi des conséquences effectives » et « loi des conséquences simulées ».

Plusieurs décennies d'expériences sur l'apprentissage conduites sur des animaux montrent que les actions qui ont le plus rapporté par le passé sont plus fréquemment choisies. Les psychologues comportementalistes appelle cela la « loi des conséquences » (Thorndike, 1911; Herrnstein, 1970). Si on se resitue dans le cadre du modèle EWA, ceci correspond à la loi des conséquences effective. Il se trouve en fait que pendant plusieurs années, les comportementalistes estimaient que seules les récompenses pour les choix effectifs produisaient des conséquences sur le

15. Camerer et Ho (1998) montrent que le modèle EWA permet une meilleure adéquation à des données de jeux de coordination qu'une combinaison convexe des modèles renforcement-croyances.

comportement et négligeaient en fait les constructions mentalistes, comme l'imagination, qui permettent aux agents de tenir compte des récompenses hypothétiques dans le processus de choix de nouvelles actions. Cette frontière s'estompa cependant lorsque plusieurs séries d'expériences mirent en évidence la pertinence des constructions cognitives dans le comportement des individus. Un concept supplémentaire est ici introduit, la loi des conséquences simulées, qui énonce que les actions non choisies qui auraient pu produire des paiements élevés – les succès simulés – sont ensuite choisies avec une plus grande probabilité. Ce nouvel aspect fournit une vision plus concrète de ce que l'on entend en général lorsque l'on parle d'apprentissage. Par exemple, lorsque l'on évoque l'apprentissage pour des machines, et même parfois dans le cadre d'apprentissage humain, il est courant de supposer que la dynamique d'apprentissage est guidée, non pas par un processus de renforcement, mais par un processus de réduction des erreurs. Or, puisque les erreurs se mesurent comme la différence entre ce qui a été obtenu et ce qui aurait pu être obtenu, elles prennent bien en considération à la fois les paiements actuels et hypothétiques.

Séparer les effets empiriques de la loi des conséquences effectives et de la loi des conséquences simulées représente la clé pour distinguer différents modèles d'apprentissage. Dans le modèle EWA, la force relative de ces deux effets est donc calibrée par le paramètre δ. Le renforcement pur implique que seuls les conséquences effectives comptent ($\delta = 0$), alors que les modèles de croyances font implicitement l'hypothèse que les conséquences effectives et simulées ont la même force ($\delta = 1$). Le modèle de EWA place le curseur entre ces deux extrêmes.

Ce modèle a été utilisé pour expliquer et prédire les choix des individus dans une large gamme de jeux : des jeux de coordination (Camerer et Ho, 1999) aux jeux à stratégies mixtes (Camerer et Ho, 1999), en passant par les concours de beauté (Camerer et Ho, 1999), les jeux de signaux (Anderson et Camerer, 2000) ou encore les jeux du Centipede (Ho *et al*., 2008) parmi d'autres. Les valeurs de δ tendent, de manière robuste, à se situer entre 0,5 et 1 dans la plupart des études, à l'exception de celles basées sur des jeux ne possédant que des équilibres en stratégies mixtes, pour lesquels δ est proche de 0. Ceci confirme que, cette dernière classe de jeux mise à part, l'approche par les croyances est plus pertinente que le simple renforcement pavlovien. Les valeurs de ϕ se situe quant à elles autour de 0,9, ce qui signifie que selon les postulats du modèle EWA, il existe une inertie relativement forte concernant les attractions passées. Enfin, les valeurs de κ sont généralement assez faibles et le plus souvent proches de 0, indiquant que les attractions croissent à un faible taux.

Deux principales critiques ont été émises par les chercheurs à l'égard du modèle EWA. Premièrement, ce modèle a plusieurs paramètres à estimer et il convient alors de se demander si l'adéquation du modèle aux données n'est pas surévaluée.

Une seconde critique à laquelle s'est heurtée l'approche EWA est que les valeurs des paramètres étaient susceptibles de varier d'un jeu à l'autre (bien que ceci soit valable pour tous les modèles d'apprentissage, voir Cheung et Friedman, 1997 ou Erev et Roth, 1998). Ce qui veut dire que fournir des prédictions sur des jeux différents nécessiterait de considérer une fonction des jeux vers les valeurs de paramètres.

Pour pallier à ces deux limites, Ho *et al.* (2007) ont développé une variante du modèle EWA, le EWA auto-ajusté. Ce faisant, ils essaient de créer une théorie adéquate avec seulement un paramètre libre : la sensibilité des joueurs à leurs attractions (qui peut d'ailleurs être abandonné au profit d'un processus de meilleure réponse déterministe si l'objectif est simplement de maximiser le *hit rate*, c'est-à-dire la fréquence avec laquelle l'action prédite est choisie par le joueur). Plus précisément, dans cette nouvelle approche, les auteurs remplacent certains paramètres du modèle EWA par des formes fonctionnelles ou des valeurs plausibles, de telle sorte que ces paramètres n'ont plus besoin d'être estimés.

2.4 *Vers une approche plus sophistiquée du comportement des joueurs*

Une caractéristique commune à toutes les approches développées ci-dessus est qu'elle considère les joueurs comme totalement adaptatifs, ne réalisant pas que leurs actions courantes sont en mesure d'influencer leurs opposants dans le futur. Ces joueurs présument en fait que le comportement de leurs opposants est généré par un processus exogène. En d'autres termes, selon ces approches, les joueurs négligent les interactions stratégiques dans leurs jeux. Bien évidemment, ce postulat est, on le comprend aisément, susceptible de sonner faux à l'oreille de bon nombre de théoriciens des jeux et il est naturel d'examiner la validité de cette hypothèse *a priori* forte. Il est certain que les modèles d'apprentissage se sont montrés efficaces pour décrire les comportements dans une classe de jeux très large et très diversifiée, ce qui souligne sans aucun doute la valeur ajoutée de ces approches dans l'étude du comportement des joueurs. Cependant, des questions restent en suspens. Est-ce que des approches plus en lien avec les fondements de la théorie des jeux sont possibles et si oui, serait-elles en mesure d'apporter un supplément d'information dans la description du comportement des joueurs ? De manière équivalente, on pourrait être tenté d'étendre le champ de l'apprentissage en introduisant de la sophistication dans le raisonnement des joueurs et voir si cela aide à rendre compte des données avec une plus grande acuité.

En particulier, les joueurs sophistiqués peuvent être conscients que leurs opposants apprennent et peuvent utiliser cette connaissance du processus d'apprentissage des autres pour en tirer profit et influencer le comportement de leurs opposants par leurs actions. La profondeur de raisonnement des joueurs sophistiqués peut ainsi leur permettre d'anticiper le comportement de leurs opposants et peuvent typiquement choisir des actions sous optimales à court terme, mais qui sont susceptibles d'influencer leurs opposants et de conduire à des paiements plus importants dans le long terme. À l'instar de Camerer *et al.* (2002a), nous parlerons pour la suite de « manipulation stratégique » pour désigner l'utilisation de ces stratégies de jeu répété à l'encontre d'agents adaptatifs, ajustant leur comportement de manière passive et myope en réponse à l'histoire du jeu. Cette manipulation stratégique s'apparente en réalité à un apprentissage d'ordre supérieur, de l'apprentissage sur l'apprentissage des opposants ou encore une appréhension plus ou moins correcte de la manière avec laquelle les autres apprennent dans leurs jeux.

Au-delà de l'intuition raisonnable qui amène à penser que les individus sont capables de manipuler les autres dans des environnements stratégiques, les stratégies sophistiquées sont en mesure de rationaliser un certains nombre de situations. En réalité, la manipulation stratégique se rencontre depuis toujours et dans un grand nombre de domaines de l'économie. Par exemple, les analystes de Wall Street tendent à faire baisser le cours des actions des compagnies qui rapportent des rendements décevants en deça de ce qu'ils attendaient. Les gestionnaires ont alors des incitations à manipuler les anticipations sur les rendements autant qu'ils le peuvent. Ils ont donc intérêt à manipuler à la baisse les anticipations sur les rendements (en utilisant par exemple des méthodes comptables pour dissimuler une part des rendements potentiels) de manière à ce que des rendements plus élevés que ceux attendus créent une surprise positive pour les analystes.

D'autre part, dans le processus de manipulation stratégique, il est important de comprendre la manière avec laquelle un rival apprend de façon à éviter de trop manipuler ou pas suffisamment. Prenons le cas d'un fournisseur qui tente de rassurer un nouveau client sur sa bonne volonté en offrant des biens supplémentaires, des garanties à moindre prix, *etc.* Certains clients apprendront facilement que le fournisseur est fiable, de sorte que répéter les concessions ou les extras est en fait une perte de revenus (le fournisseur manipule de manière excessive). D'autres clients, plus méfiants, accorderons plus difficilement leur confiance et le fournisseur aura besoin de multiplier les gestes commerciaux avant que la relation de confiance ne s'installe de manière durable.

Il existe une grande quantité d'applications importantes du concept de manipulation stratégique. Dans le champ de l'économie du travail, il est bien établi qu'une poussée d'inflation non anticipée peut faire baisser le niveau de l'emploi (c'est ce qu'implique la courbe de Philips). Ainsi, des décideurs possédant une optique de long terme, comme par exemple les dirigeants de la Banque centrale européenne ou de de la Réserve fédérale américaine, souhaiteront donner confiance aux agents en maintenant le niveau d'inflation bas pendant une certain temps, ce qui permet de certifier leur capacité à résister aux tentations inflationnistes. Autrement dit, le décideur a une incitation à manipuler les agents en leur laissant entendre que le niveau d'inflation restera faible dans le futur. Cette idée a été utilisée pour expliquer le passage d'un taux d'inflation à deux chiffres dans les années 70 aux États-Unis à une chute de l'inflation dans les années ultérieurs à 1985[16].

Un autre exemple concerne le détermination du prix d'un bien sur des marchés très volatils tels que les ordinateurs, les téléphones intelligents ou autres produits de nouvelle technologie. Sur ces marchés, les consommateurs sont susceptibles de reporter leurs consommations s'ils anticipent que les prix vont baisser rapidement. Dès que les firmes préfèrent vendre leur bien plus tôt que plus tard, elles ont alors une incitation à manipuler les consommateurs en leur faisant croire que les prix resteront élevés pendant un certains temps, de telle sorte qu'ils achèteront le produit sans attendre.

16. Voir Sargent (1999).

Toujours dans le domaine de l'économie industrielle, les firmes peuvent être incitées à se manipuler les unes les autres. Plus précisément, les firmes peuvent se comporter de manière agressive et produire des quantités élevées de biens pour induire un comportement futur plus passif de leurs rivaux et également dissuader de nouveaux concurrents à entrer dans le marché.

Tous les exemples ci-dessus suggèrent que l'étude de l'apprentissage n'est pas suffisante. Pour se faire une idée plus précise de la façon avec laquelle les agents interagissent à long terme, il est nécessaire de prendre en compte le fait que les joueurs sont capables de réaliser que les autres sont effectivement en mesure d'apprendre et de répondre de leur expérience récente. Ainsi les individus peuvent retourner cette expérience à leur avantage. En résumé, les joueurs sont susceptibles d'être plus sophistiqués que ce que les modèles d'apprentissage classiques supposent.

Une première recherche de Camerer *et al.* (2002a) introduit de la sophistication dans les modèles d'apprentissage adaptatifs en considérant une population composée de deux types d'agents : une fraction d'entre eux est supposée parfaitement rationnels et peut par conséquent exhiber le sentier d'équilibre quand la fraction résiduelle des joueurs est supposée adaptative et n'actualise ses actions présentes qu'au regard de l'expérience passée. Les joueurs rationnels ont à l'esprit leur propre répartition de la population (pas nécessairement la vraie) entre agents rationnels et adaptatifs et utilisent cette connaissance pour anticiper le comportement des individus adaptatifs. En somme, ces auteurs examinent l'apprentissage dans une population hétérogène de joueurs. Ils estiment leur modèle sur la base de données expérimentales de concours de beauté et de jeux d'investissements répétés. Leur modèle de manipulation stratégique s'avère meilleur que le modèle EWA pour décrire et prédire les comportements.

Cependant, l'approche de Camerer *et al.* (2002a) possède notamment une limite importante. Selon ce modèle, les individus rationnels ne sont jamais amenés à réviser leur perception initiale du processus d'apprentissage de leurs opposants. Par conséquent, ceux-ci ne sont pas censés arrêter des tentatives de manipulation infructueuses ou à l'inverse se lancer dans des stratégies de manipulation en cours de jeu. Si cette hypothèse peut sembler plausible dans des populations larges où les joueurs sont appariés de manière aléatoire de période en période (comme c'est le cas dans les expériences utilisées dans Camerer *et al*., 2002a), elle apparaît beaucoup plus critiquable au sein de populations où la fréquence des interactions sociales est plus forte. Cette limite est particulièrement problématique dans la mesure où un certain nombre d'études récentes témoignent notamment de l'usage de stratégies non stationnaires dans les jeux répétés (voir entre autres Terracol et Vaksmann, 2009; Hyndman *et al.,* 2009; Hyndman *et al*., 2012). Dans leur étude, Hyndman *et al.* (2009) introduisent un concept d'« apprentissage sophistiqué », ou « apprentissage de second ordre » selon lequel les agents perçoivent les conséquences futures de leurs propres actions mais conservent un degré de rationalité limitée dans la mesure où ils réevaluent de manière adaptative le processus d'apprentissage de leurs opposants en fonction des comportements observés dans le passé. Les joueurs

sophistiqués ajustent notamment leur perception de la vitesse d'apprentissage de leurs opposants selon le niveau d'inertie dont ceux-ci ont fait preuve dans le choix de leurs actions passées. Cette approche permet en outre d'expliquer de manière plus satisfaisante la dynamique de la manipulation stratégique et a plus spécifiquement prouvé son efficacité par sa capacité à rendre compte avec acuité des comportements observés notamment dans plusieurs variantes de jeux de coordination.

2.5 *Remarques conclusives*

Le processus par lequel un équilibre émerge a été largement ignoré jusqu'à la fin du XX^{ème} siècle. Mais l'essor de modèles capables de décrire le raisonnement des joueurs permet maintenant d'y voir plus clair sur cette question. Plus récemment, les recherches se sont penchées sur un processus d'équilibration mettant en actions des agents complètement myopes/adaptatifs. Les économistes ont dans un premier temps suivi les traces des psychologues en reprenant leurs approches d'apprentissage par renforcement. L'idée du renforcement est que les joueurs tendent à manifester une plus forte propension à choisir des actions qui ont rapporté davantage dans le passé.

Cependant, cette approche est apparue extrêmement limitée pour décrire le raisonnement humain puisque des agents intelligents sont typiquement susceptibles de baser leurs choix sur des motivations plus élaborées. Ainsi, un modèle nouveau, basé sur des considérations plus proches des fondements de la théorie des jeux, a été mis au point. Plus précisément, les économistes ont proposé des théories d'apprentissage par les croyances selon lesquelles les agents ont typiquement tendance à opter pour les actions qui leur procurent l'espérance de gains la plus importante étant donné les croyances qu'ils se forment. Ceci a amené directement à la question du processus de formation des croyances individuelles. Les premiers procédés heuristiques utilisés ont été ceux de Cournot et de *fictitious play.* Dans le premier cas, les agents attribuent une probabilité égale à 1 à l'action précédemment jouée par leurs adversaires. Dans le second cas, la croyance concernant une action donnée correspond à la fréquence avec laquelle cette action a été choisie sur l'ensemble de l'historique disponible. Cheung et Friedman (1997) ont fourni par la suite un modèle empirique plus général de formation de croyances, le *fictitious play* pondéré, qui incorpore comme cas particuliers les approches de Cournot et le *fictitious play* traditionnel. De manière générale, les modèles d'apprentissage par les croyances se sont montrés relativement performants pour retranscrire de manière fidèle le comportement des joueurs, à l'exception des jeux ne possédant que des équilibres en stratégies mixtes (non dégénérées).

Les apprentissages par renforcement et par croyances ont été considérés complètement distincts jusqu'à la fin des années 90 et aux travaux de Camerer et Ho (1999). Ces auteurs ont conçu un modèle hybride d'apprentissage EWA qui inclut l'apprentissage par renforcement et celui par les croyances comme cas spéciaux. Ce modèle EWA a fait ses preuves dans un grand nombre de contextes mais a en même temps été considéré comme techniquement complexe à mettre en application en raison de son grand nombre de paramètres libres à estimer. Ainsi, Ho *et al.* (2007)

ont proposé une variante à un paramètre du modèle EWA, le modèle EWA auto-ajusté, dans lequel la plupart des paramètres sont remplacés par des règles empiriques ou des valeurs plausibles de telle sorte qu'ils n'ont plus besoin d'être estimés.

Cependant selon toutes ces approches, les joueurs adoptent un comportement relativement primitif et prennent leurs décisions sans considérer les interactions stratégiques. Plus spécifiquement, les joueurs choisissent leurs actions sur la base de ce qu'ils ont expérimenté dans le passé et ne réalisent pas que leurs opposants sont également en mesure d'apprendre. Ainsi ils ne perçoivent pas que leurs propres actions peuvent influencer le comportement de leurs opposants dans le futur. Il va sans dire que ce postulat peut apparaître déconnecté des fondements de la théorie des jeux et la recherche d'approches plus sophistiquées se fait alors ressentir. Dans cette optique Camerer *et al.* (2002a) propose un modèle de manipulation stratégique dans lequel les joueurs peuvent être, soit purement adaptatifs comme le supposent l'ensemble des théories de l'apprentissage, soit parfaitement rationnels. Ces derniers détiennent notamment la capacité cognitive d'anticiper le comportement futur de leurs opposants sur la base de l'histoire passée du jeu. Ils peuvent donc tirer profit des interactions stratégiques et manipuler leurs opposants passifs. Cette approche présente ainsi l'intérêt d'intégrer des comportements plus sophistiqués dans les modèles d'apprentissage passif usuels. Néanmoins, selon ce modèle les agents rationnels ne sont jamais amenés à réviser leurs *a priori* sur le processus d'apprentissage de leurs opposants ou sur la fraction d'individus passifs dans la population. Cette hypothèse est en outre incompatible avec certaines régularités observées dans une classe élargie de jeux qui montrent que les joueurs sont par exemple susceptibles de cesser à un moment donné d'utiliser des stratégies de manipulation en réalisant que la vitesse d'apprentissage de leurs opposants est plus lente qu'ils ne l'avaient initialement escomptée. Dans une autre approche, Hyndman *et al.* (2009) introduisent un modèle dans lequel les agents possèdent un degré de sophistication stratégique supérieur à celui supposé dans les modèles d'apprentissage passif mais conservent par ailleurs un degré d'adaptativité en réévaluant leur perception du processus d'apprentissage de leurs opposants sur la base des comportements observés. Cette approche, qualifiée d'apprentissage sophistiqué, ou d'apprentissage de second ordre, permet notamment de rationaliser les régularités évoquées plus haut quant à la dynamique de la manipulation stratégique. Elle s'est en outre montrée efficace pour ce qui était de retracer le comportement des joueurs dans plusieurs variantes de jeux de coordination.

Conclusion

Cet article a passé en revue un certain nombre de modèles visant à reproduire les déviations observées de l'équilibre de Nash et à fournir une explication cognitivement réaliste à ces derniers. Une première branche de cette littérature s'intéresse aux réactions initiales des agents, sans expérience passée du jeu.

Un second pan de la littérature étudie la dynamique du comportement des agents dans les jeux répétés et la manière avec laquelle ceux-ci utilisent leur expérience passée pour actualiser leurs actions courantes.

Plus généralement, les modèles de rationalité limitée, initialement conçus dans le but de relâcher la forte complexité cognitive présente dans les concepts pionniers de la théorie des jeux, prennent depuis quelques années une voie inverse en s'attachant à fournir de nouvelles élaborations plus sophistiquées pour dépasser les limites des modèles comportementaux de première vague. La ligne directrice des travaux, passés et à venir, dans ce domaine restent dictée par des standards empiriques et la volonté de produire des théories conformes au niveau de sophistication stratégique observé chez les agents.

BIBLIOGRAPHIE

AGRANOV, M., A. CAPLIN et C. TERGIMAN (2015), « Naive Play and the Process of Choice in Guessing Games » *Journal of the Economic Science Association*, 1(2): 146-157

AGRANOV, M., E. POTAMITES, A. SCHOTTER et C. TERGIMAN (2012), « Beliefs and Endogenous Cognitive Levels: An Experimental Study », *Games and Economic Behavior*, 75(2) : 449 463.

ANDERSON, C. M. et C. F. CAMERER (2000), « Experience-Weighted Attraction Learning in Sender-Receiver Signaling Games », *Economic Theory*, 16(3) : 689-718.

ANDERSON, S. P., J. K. GOEREE et C. A. HOLT (1998), « Rent Seeking with Bounded Rationality: An Analysis of the All-Pay Auction », *Journal of Political Economy*, 106 (4) : 828-853.

ANDERSON, S. P., J. K. GOEREE et C. A. HOLT (2001), « Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium », *Games and Economic Behavior*, 34 (2) : 177-199.

ARAD, A. et A. RUBINSTEIN (2012), « The 11-20 Money Request Game: A Level-k Reasoning Study », *American Economic Review*, 102(7) : 3561-3573.

ARTHUR, W. B. (1991), « Designing Economic Agents that Act Like Human Agents: A Behavioral Approach to Bounded Rationality », *American Economic Review*, 81(2) : 353-359.

ARTHUR, W. B. (1994). « On Designing Economic Agents that Behave like Human Agents », *Journal of Evolutionary Economics*, 3 : 1-22.

AUMANN, R. et A. BRANDENBURGER (1995), « Epistemic Conditions for Nash Equilibrium », *Econometrica*, 63(5) : 1161-1180.

BATTALIO, R., L. SAMUELSON et J. VAN HUYCK (2001), « Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games », *Econometrica*, 69(3) : 749-764.

BELLHOUSE, D. (2007), « The Problem of Waldegrave », *Journal Électronique d'Histoire des Probabilités et de la Statistique,* 3(2).

BERNHEIM, B. D. (1984), « Rationalizable Strategic Behavior », *Econometrica*, 52(4) : 1007-1028.

BOYLAN, R. T. et M. A. EL-GAMAL (1993), « Fictitious Play: A Statistical Study of Multiple Economic Experiments », *Games and Economic Behavior*, 5(2) : 205-222.

BRANDTS, J. et C. HOLT (1996), « Naive Bayesian Learning and Adjustment to Equilibrium in Signaling Games », University of Virginia, Discussion Paper.

BROWN, G. (1951), « Iterative Solution of Games by Fictitious Play », *in* T. C. KOOPMANS (éd.), *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, New York : Wiley, 404 p.

BURCHARDI, K. B. et S. P. PENCZYNSKI (2014), « Out of your Mind: Eliciting Individual Reasoning in one Shot Games », *Games and Economic Behavior*, 84(C) : 39-57.

BURNHAM, T. C., D. CESARINI, M. JOHANNESSON, P. LICHTENSTEIN et B. WALLACE (2009), « Higher Cognitive Ability is Associated with Lower Entries in a p-Beauty Contest », *Journal of Economic Behavior and Organization*, 72(1) : 171-175.

BUSH, R. R. et F. MOSTELLER (1955), *Stochastic Models for Learning*, New York : Wiley.

CAMERER, C. et T.-H. HO (1998), « EWA Learning in Coordination Games: Probability Rules, Heterogeneity, and Time Variation », *Journal of Mathematical Psychology*, 42 : 305-326.

CAMERER, C. et T.-H. HO (1999), « Experience-weighted Attraction Learning in Normal Form Games », *Econometrica*, 67(4) : 827-874.

CAMERER, C., T.-H. HO et J.-K. CHONG (2002a), « Sophisticated EWA Learning and Strategic Teaching in Repeated Games », *Journal of Economic Theory*, 104 : 137-188.

CAMERER, C. et R. KARJALAINEN (1994), « Ambiguity-Aversion and non-Additive Beliefs in non-Cooperative Games: Experimental Evidence », *in* B. MUNIER et M. MACHINA (éds), *Models and Experiments on Risk and Rationality*, Dordrecht : Kluwer, p. 325-358.

CAMERER, C. F. et K. WEIGELT (1988), « Experimental Tests of a Sequential Equilibrium Reputation Model », *Econometrica*, 56(1) : 1-36.

CAMERER, C.F., T.-H. HO et J.-K. CHONG (2002b), « Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games », *Journal of Economic Theory*, 104(1) : 137-188.

CAMERER, C. F., T.-H. HO et J. K. CHONG (2004a), « Behavioral Game Theory: Thinking, Learning, and Teaching » *in* S. HUCK (éd), *Advances in Understanding Strategic Behavior,* Essays in Honor of Werner Guth, London : Palgrave MacMillan, p. 120-180.

CAMERER, C. F., T.-H. HO et J. K. CHONG (2004b), « A Cognitive Hierarchy Model of Games », *The Quarterly Journal of Economics*, 119(3) : 861-898.

CAPRA, C. M. (1999), « Anomalous Behavior in a Traveler's Dilemma? », *American Economic Review*, 89(3) : 678-690.

CHEN, H.-C., J. W. FRIEDMAN et J.-F. THISSE (1997), « Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach », *Games and Economic Behavior*, 18(1) : 32-54.

CHEUNG, Y.-W. et D. FRIEDMAN (1997), « Individual Learning in Normal Form Games: Some Laboratory Results », *Games and Economic Behavior*, 19(1) : 46-76.

COOPER, D.J., S. GARVIN et J.H. KAGEL (1997), « Adaptive Learning vs. Equilibrium Refinements in an Entry Limit Pricing Game », *Economic Journal*, 107(442) : 553-575.

CORICELLI, G. et R. NAGEL (2009), « Neural Correlates of Depth of Strategic Reasoning in Medial Prefrontal Cortex », *Proceedings of the National Academy of Science*, 106(23) : 9163-9168.

COSTA-GOMES, M., V. P. CRAWFORD et B. BROSETA (2001), « Cognition and Behavior in Normal-Form Games: An Experimental Study », *Econometrica*, 69(5) : 1193-1235.

COSTA-GOMES, M. A. et V. P. CRAWFORD (2006), « Cognition and Behavior in Two-Person Guessing Games : An Experimental Study », *American Economic Review*, 96(5) : 1737-1768.

COSTA-GOMES, M. A., V. P. CRAWFORD et N. IRIBERRI (2009), « Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games », *Journal of the European Economic Association*, 7(2-3) : 365-376.

COSTA-GOMES, M. A. et G. WEIZSÄCKER (2008), « Stated Beliefs and Play in Normal-Form Games », *Review of Economic Studies*, 75(3) : 729-762.

COURNOT, A. (1838), *Recherches sur les principes mathématiques de la théorie des richesses*, translated into english by N. BACON (1960) as *Researches in the Mathematical Principles of the Theory of the Wealth*, London : Haffner.

CRAWFORD, V. P., M.A. COSTA-GOMES et N. IRIBERRI (2013), « Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications », *Journal of Economic Literature*, 51(1) : 5-62.

CRAWFORD, V. P. et N. IRIBERRI (2007), « Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental "Hide-and-Seek" Games », *American Economic Review*, 97(5) : 1731-1750.

CROSS, J. G. (1973), « A Stochastic Learning Model of Economic Behavior », *The Quaterly Journal of Economics*, 87(2) : 239-266.

CROSS, J. G. (1983), *A Theory of Adaptive Economic Behavior*, NewYork/London : Cambridge University Press.

EREV, I. et A. E. ROTH (1998), « Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria », *American Economic Review*, 88(4) : 848-81.

FLOOD, M. M. (1952), « Some Experimental Games », Research Memorandum RM-789, RAND Corporation.

FLOOD, M. M. (1958), « Some Experimental Games », *Management Science*, 5(1) : 5-26.

FOSTER, D. P. et R. V. VOHRA (1997), « Calibrated Learning and Correlated Equilibrium », *Games and Economic Behavior*, 21(1-2) : 40-55.

FUDENBERG, D. et D. M. KREPS (1993), « Learning Mixed Equilibria », *Games and Economic Behavior*, 5(3) : 320-367.

FUDENBERG, D. et D. M. KREPS (1995), « Learning in Extensive-Form Games I. Self-confirming Equilibria », *Games and Economic Behavior*, 8(1) : 20-55.

FUDENBERG, D. et D. LEVINE (1998), *The Theory of Learning in Games*, MIT Press, MA : Cambridge.

BRAÑAS GARZA, P., T. GARCÍA-MUÑOZ et R. H. GONZÁLEZ (2012), « Cognitive Effort in the Beauty Contest Game », *Journal of Economic Behavior and Organization*, 83(2) : 254-260.

GOEREE, J., C. HOLT et T. PALFREY (2005), « Regular Quantal Response Equilibrium », *Experimental Economics*, 8(4) : 347-367.

GOEREE, J. K. et C. A. HOLT (2001), « Ten Little Treasures of Game Theory and Ten Intuitive Contradictions », *American Economic Review*, 91(5) : 1402-1422.

GOEREE, J. K. et C. A. HOLT (2004), « A Model of Noisy Introspection », *Games and Economic Behavior*, 46 (2) : 365-382.

GOEREE, J. K. et C. A. HOLT (2005), « An Explanation of Anomalous Behavior in Models of Political Participation », *The American Political Science Review*, 99(2) : 201-213.

GOEREE, J. K., C. A. HOLT et T. R. PALFREY (2002), « Quantal Response Equilibrium and Overbidding in Private-Value Auctions », *Journal of Economic Theory*, 104(1) : 247-272.

HAILE, P. A., A. HORTAÇSU et G. KOSENOK (2008), « On the Empirical Content of Quantal Response Equilibrium », *American Economic Review*, 98(1) : 180-200.

HARLEY, C. (1981), « Learning in Evolutionary Stable Strategies », *Journal of Theoretical Biology*, 13 : 611-633.

HERRNSTEIN, J. R. (1970), « On the Law of Effect », *Journal of Experimental Analysis of Behavior*, 13 : 342-366.

HO, T. H., C. F. CAMERER et J.-K. CHONG (2007), « Self-Tuning Experience Weighted Attraction Learning in Games », *Journal of Economic Theory*, 133(1) : 177-198.

HO, T.-H., X. WANG ET C. CAMERER (2008), « Individual Differences in the EWA Learning with Partial Payoff Information », *The Economic Journal*, 118 : 37-59.

HO, T.-H. et K. WEIGELT (1996), « Task Complexity, Equilibrium Selection, and Learning: An Experimental Study », *Management Science*, 42(5) : 659-679.

HYNDMAN, K., E. Y. OZBAY, A. SCHOTTER et W. Z. EHRBLATT (2012), « Convergence: An Experimental Study Of Teaching And Learning In Repeated Games », *Journal of the European Economic Association*, 10(3) : 573-604.

HYNDMAN, K., A. TERRACOL et J. VAKSMANN (2009), « Learning and Sophistication in Coordination Games », *Experimental Economics*, 12(4) : 450-472.

IVANOV, A., D. LEVIN et M. NIEDERLE (2010), « Can Relaxation of Beliefs Rationalize the Winner's Curse?: An Experimental Study », *Econometrica*, 78(4) : 1435-1452.

KEYNES, J. M. (1936), *The General Theory of Employment, Interest and Money*, New York : Harcourt Brace and Co.

Mc Allister, P. (1991), « Adaptive Approaches to Stochastic Programming », *Annals of Operations Research*, 30 : 45-62.

McKelvey, D. R. et R. Page (1990), « Public and Private Information: an Experimental Study of Information Pooling », *Econometrica*, 58 : 1321-1339.

McKelvey, D. R. et T. R. Palfrey (1995), « Quantal Response Equilibria for Normal Form Games », *Games and Economic Behavior*, 10(1) : 6-38.

McKelvey, D. R. et T. R. Palfrey (1998), « Quantal Response Equilibria for Extensive Form Games », *Experimental Economics*, 1 : 9-41.

Moinas, S. et S. Pouget (2013). « The Bubble Game: An Experimental Study of Speculation », *Econometrica*, 81(4) : 1507-1539.

Mookherjee, D. et B. Sopher (1994), « Learning Behavior in an Experimental Matching Pennies Game », *Games and Economic Behavior*, 7(1) : 62-91.

Mookherjee, D. et B. Sopher (1997), « Learning and Decision Costs in Experimental Constant Sum Games », *Games and Economic Behavior*, 19(1) : 97-132.

Nagel, R. (1995), « Unraveling in Guessing Games: An Experimental Study », *American Economic Review*, 85(5) : 1313-1326.

Nash, J. F. (1950), « Equilibrium Points in n-Person Games », *Proceedings of the National Academy of Science*, 36(1) : 48-49.

von Neumann, J. (1928), « Zur Theorie der Gesellschaftsspiele », *Mathematische Annalen*, 100(1) : 295-320.

von Neumann, J. et O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton University Press.

Nyarko, Y. et A. Schotter (2002), « An Experimental Study of Belief Learning Using Elicited Beliefs », *Econometrica*, 70(3) : 971-1005.

Ostling, R., J. T. Wang, E. Y. Chou et C.F. Camerer, (2011), « Testing Game Theory in the Field: Swedish LUPI Lottery Games », *American Economic Journal: Microeconomics*, 3(3): 1-33.

Pearce, D. G. (1984), « Rationalizable Strategic Behavior and the Problem of Perfection », *Econometrica*, 52(4) : 1029-50.

Polak, B. (1999), « Epistemic Conditions for Nash Equilibrium, and Common Knowledge of Rationality », *Econometrica*, 67(3) : 673-676.

Robinson, J. (1951), « An Iterative Method of Solving a Game », *Annals of Mathematics*, 54 : 296-301.

Roth, A. E. et I. Erev (1995), « Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term », *Games and Economic Behavior*, 8(1) : 164-212.

Sargent, T. J. (1999), « A Primer on Monetary and Fiscal Policy », *Journal of Banking and Finance*, 23(10) : 1463-1482.

Sarin, R. et F. Vahid (1999), « Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice », *Games and Economic Behavior*, 28(2) : 294-309.

SARIN, R. et F. VAHID (2001), « Predicting How People Play Games: A Simple Dynamic Model of Choice », *Games and Economic Behavior*, 34(1) : 104-122.

SELTEN, R. (1991), « Evolution, Learning, and Economic Behavior », *Games and Economic Behavior*, 3(1) : 3-24.

SHAPLEY, L. (1964), « Some Topics in Two-Person Games » *in* M. Dresher, L.S. Shapley et A.W. Tucker (éds), *Advances in Game Theory*, Annals of Mathematical Studies 52, Princeton University Press, p. 1-28.

SIMON, H. A. (1955), « A Behavioral Model of Rational Choice », *The Quarterly Journal of Economics*, 64(1) : 99-118.

STAHL, D. O. et P. W. WILSON (1994), « Experimental Evidence on Players' Models of other Players », *Journal of Economic Behavior and Organization*, 25(3) : 309-327.

STAHL, D. O. et P. W. WILSON (1995), « On Players' Models of Other Players : Theory and Experimental Evidence », *Games and Economic Behavior*, 10(1) : 218-254.

TERRACOL, A. et J. VAKSMANN (2009), « Dumbing Down Rational Players: Learning and Teaching in an Experimental Game », *Journal of Economic Behavior and Organization*, 70(1-2) : 54-71.

THORNDIKE, E. L. (1911), *Animal intelligence*, New York : Macmillan.

WEIZSÄCKER, G. (2003), « Ignoring the Rationality of Others : Evidence from Experimental Normal-Form Games », *Games and Economic Behavior*, 44(1) : 145-171.