Revue d'économie industrielle

114-115 | 2e-3e trimestre 2006
Processus de contagion et interactions stratégiques

Comportements adaptatifs sans croyances

Sylvain Béal et Jacques Durieu

Édition électronique
URL : http://journals.openedition.org/rei/358
DOI : 10.4000/rei.358
ISSN : 1773-0198

Éditeur
De Boeck Supérieur

Édition imprimée
Date de publication : 15 septembre 2006
Pagination : 119-140
ISSN : 0154-3229

Référence électronique
Sylvain Béal et Jacques Durieu, « Comportements adaptatifs sans croyances », *Revue d'économie industrielle* [En ligne], 114-115 | 2e-3e trimestre 2006, mis en ligne le 29 novembre 2007, consulté le 19 avril 2019. URL : http://journals.openedition.org/rei/358 ; DOI : 10.4000/rei.358

Sylvain BEAL, Jacques DURIEU
CREUSET, Université de Saint-Étienne

COMPORTEMENTS ADAPTATIFS SANS CROYANCES

Mots-clés : Adaptation, renforcement, seuil d'aspiration, imitation.

Key words : Evolution, Fictitious Play, Imitation, Game Theory, Stochastic Stability.

I. — INTRODUCTION

Dans son acception standard, la théorie des jeux non coopératifs traite de situations d'interaction stratégique entre des joueurs qui sont pleinement rationnels. Cette approche s'appuie notamment sur des hypothèses fortes en termes de connaissance de la structure du jeu. Parallèlement, une approche complémentaire se développe depuis une vingtaine d'années en modélisant des situations d'interaction stratégique prenant place entre des agents munis d'une rationalité et de connaissances limitées. À l'inverse de la théorie des jeux standard, cette approche retient une hypothèse d'information structurelle incomplète. En d'autres termes, chaque agent peut avoir une information plus ou moins complète sur sa propre fonction de paiement ainsi que sur les caractéristiques de ses opposants. Chaque agent a donc une compréhension incomplète de son environnement. En outre, les capacités cognitives dont sont dotés les individus sont limitées et sont susceptibles d'entraver la résolution de problèmes d'optimisation complexes. En conséquence, les agents tentent de simplifier leur tâche de décision. Cette simplification peut être accomplie à deux niveaux. Premièrement, les individus peuvent opter pour des représentations élémentaires de leur environnement. Une hypothèse traditionnelle consiste ainsi à retenir une situation de récurrence d'un jeu donné et à supposer que chaque agent croit en la stationnarité de l'environnement d'une période sur l'autre. Cette croyance signifie que les agents ne perçoivent pas que leurs opposants sont engagés dans une procédure de révision de leur stratégie. Ils ne comprennent donc pas que leurs propres choix peuvent modifier les actions jouées par leurs opposants. Deuxièmement, les individus adoptent des procédures de décision peu sophistiquées, leur prescrivant un comportement à tenir

pour la période à venir. En contrepartie, les agents ont la possibilité d'exploiter les informations factuelles collectées au long de l'histoire constituée par la récurrence des situations d'interaction. Les procédures de choix des stratégies se caractérisent par un principe d'adaptation compte tenu des informations factuelles rétrospectives considérées.

Les modèles basés sur de tels comportements adaptatifs adressent alors la question de la convergence temporelle des profils de stratégies jouées. Particulièrement, il s'agit de déterminer sous quelles conditions les processus évolutionnaires peuvent converger vers les équilibres de Nash du jeu récurrent. Autrement dit, il s'agit d'évaluer si l'information factuelle est susceptible de pallier le manque d'information structurelle et la rationalité limitée des agents. Cependant, dans cette catégorie de modèles, l'information factuelle ne peut en aucun cas servir à reconstituer l'information structurelle. Les individus n'apprennent pas à comprendre et à former une représentation correcte de leur environnement (1). L'accent est donc mis sur la capacité d'adaptation dont peuvent faire preuve les agents dans leurs choix, compte tenu d'une représentation de l'environnement donnée initialement et non remise en cause au fil du temps.

Une difficulté avec cette approche est qu'il n'existe pas un unique modèle de comportement relevant de la rationalité limitée. L'information factuelle rétrospective peut être exploitée de plusieurs manières et les procédures de choix peuvent revêtir des formes diverses. Un des objectifs de cet article, ainsi que des contributions suivantes, est de présenter des modèles étudiant certains comportements adaptatifs. Afin d'éclairer les hypothèses propres à chaque modèle et de rendre compte de la diversité des comportements appréhendés, deux critères de démarcation des règles comportementales peuvent être avancés.

Le premier critère s'appuie sur la représentation de l'environnement élaborée par les agents. Deux cas de figure sont à considérer. Premièrement, les agents ont conscience de la nature stratégique de l'environnement : ils comprennent que leurs propres paiements dépendent à la fois de leurs propres actions et des actions choisies par les opposants. En conséquence, les agents mobilisent des informations factuelles rétrospectives disponibles afin de former des croyances sur les stratégies futures de leurs opposants. Deuxièmement, les agents ne prennent pas en considération la dimension stratégique des interactions et n'élaborent pas de croyances sur les comportements futurs des autres agents. Deux facteurs sont susceptibles d'expliquer cet état de

(1) Ceci signifie que les modèles de type apprentissage bayésien dans lequel les joueurs apprennent à anticiper le comportement de leurs opposants sont exclus de cette approche.

fait. Soit les agents ne comprennent pas que leurs paiements dépendent des actions choisies par leurs opposants, soit ils appréhendent correctement la dimension stratégique de l'environnement mais ils ne l'intègrent pas dans leur prise de décision. Dans ce dernier cas, les agents peuvent ne pas disposer des informations factuelles rétrospectives concernant leurs opposants ou bien ils peuvent renoncer volontairement à former des croyances du fait de coûts jugés excessifs en termes de capacités cognitives.

Le second critère concerne les modalités d'évaluation des actions. Deux catégories sont aisément identifiables. Premièrement, les agents ont conscience de la dimension stratégique des interactions. Des informations factuelles rétrospectives concernant les stratégies jouées dans le passé par les opposants sont alors employées pour former des croyances sur le comportement futur des autres agents. Et, étant données les croyances élaborées, les agents évaluent leurs propres actions par le calcul des paiements espérés. Dans ce cas, la procédure d'évaluation des actions est donc dirigée vers le futur : un classement des actions est constitué sur la base des paiements attendus pour la période suivante. Deuxièmement, lorsque les agents n'élaborent pas de croyances sur les stratégies futures de leurs opposants, des informations factuelles rétrospectives interviennent directement pour évaluer les actions disponibles. À cette fin, les agents peuvent recourir soit à des informations concernant leurs propres paiements passés, soit à des informations collectées auprès d'autres agents. L'évaluation des actions est alors dirigée vers le passé : les actions sont valorisées en fonction de leurs performances passées.

L'objectif de ce papier est de revenir sur certains modèles dans lesquels les agents n'élaborent aucune croyance sur le comportement futur de leurs opposants (2). Les procédures d'évaluation des actions sont alors dirigées vers le passé. L'examen de ces règles comportementales est mené dans le cadre d'interactions stratégiques entre agents et, parfois, dans des jeux contre la Nature, où Nature est caractérisée par une loterie sur un ensemble d'états. Ce dernier cas de figure peut s'interpréter comme une situation d'interaction simplifiée où un agent interagit avec un opposant qui joue une stratégie mixte stationnaire dans le temps. Deux catégories de modèles retiennent notre attention. D'un côté, nous considérons des modèles où l'évaluation des actions est fondée sur les informations factuelles rétrospectives propres à chaque agent. Précisément, nous développons deux types de procédure d'évaluation des actions fondée sur des informations privées : soit chaque agent évalue les actions sur la base des paiements absolus reçus, soit chaque agent évalue les actions sur la base de paiements relatifs. Dans ce dernier cas de figure, l'agent prend en considéra-

(2) Des modèles incorporant de telles croyances sont développés dans les contributions suivantes de Baron et Solal et de Berninghaus, Haller et Outkin.

tion les paiements reçus en comparaison avec un seuil d'aspiration. Des actions jugées non satisfaisantes au regard du seuil d'aspiration se trouvent alors dévalorisées.

D'un autre côté, nous traitons de modèles où l'évaluation des actions requiert des informations factuelles rétrospectives concernant des opposants. Plusieurs modèles sont envisageables selon la nature et le volume de ces informations. Particulièrement, différentes règles comportementales sont concevables selon le nombre d'opposants et/ou le nombre de périodes passées sur lesquels l'information disponible porte. Notre propos consiste dès lors à présenter certaines de ces règles et à tenter de les justifier. Il s'agit notamment d'introduire une procédure d'évaluation portant sur les règles comportementales et permettant de décréter si une règle est satisfaisante ou non au regard d'un certain critère.

La suite de ce texte est organisée de la façon suivante. La section 2 est consacrée à la présentation de modèles incorporant une procédure d'évaluation des actions reposant sur les paiements absolus reçus dans le passé par l'agent. La section 3 traite des modèles introduisant un seuil d'aspiration dans l'évaluation des actions. La section 4 porte sur les procédures d'évaluation des actions faisant intervenir les paiements passés de certains opposants.

II. — MODÈLES DE RENFORCEMENT POSITIF

Nous considérons le cadre d'interaction stratégique suivant. Soit un jeu fini à n joueurs sous forme stratégique

$$G = (N, (X_i)_{i \in N}, (u_i)_{i \in N}).$$

L'ensemble $N = \{1, \ldots, n\}$ désigne la population finie d'agents ou de joueurs de taille n, X_i est l'ensemble fini des stratégies pures ou actions à la disposition du joueur i. Soit $X = \Pi_{i \in N} X_i$ l'ensemble des profils de stratégies pures. Pour chaque $i \in N$, notons $X_{-i} = \Pi_{j \in N \setminus \{i\}} X_j$. Soit $\Delta(X_i)$ l'ensemble des distributions de probabilité sur l'ensemble X_i. Introduisons pour chaque joueur $i \in N$ une fonction de paiement $u_i : X \to \mathbb{R}$ qui associe à chaque profil $x \in X$ un gain réel $u_i(x)$.

Supposons que $t = 1, 2, \ldots$ désigne des périodes de temps successives. Le jeu G est joué une fois à chacune des périodes. Le paiement reçu par le joueur $i \in N$ lors de la période t quand il joue l'action x_i^t et que ses opposants choisissent $x_{-i}^t \in X_{-i}$ est $u_i(x_i^t, x_{-i}^t)$. Dans un souci de simplicité, le paiement reçu par le joueur i à la période t sera parfois noté u_i^t.

Dans cette section, nous traitons de règles comportementales telles que chaque agent évalue les actions à sa disposition à partir des paiements qu'il a perçus dans le passé. En d'autres termes, seules des informations factuelles rétrospectives propres à l'agent sont mobilisées dans la procédure de choix. Les modèles fondés sur le critère de performance d'absence de regret sont

donc laissés de côté. Dans cette catégorie de modèles, les agents s'appuient sur l'histoire du jeu pour mesurer les paiements qu'ils auraient pu obtenir en effectuant des choix différents de ceux qu'ils ont effectivement faits. Ils peuvent ainsi évaluer le regret associé à chaque action et ainsi classer leurs actions (3). À l'inverse, nous mettons l'accent sur des modèles de renforcement où seuls les paiements effectivement reçus interviennent dans l'évaluation des actions.

L'idée sous-jacente à ces modèles remonte à un principe introduit par Thorndike (1898) : les actions ayant procuré de « bons » paiements dans le passé ont davantage de chance d'être utilisées dans le futur. Cette règle a été largement étudiée dans le champ de la psychologie comportementale. Ainsi, Estes (1950) et Bush et Mosteller (1955) proposent des modèles qui s'appuient sur la notion de renforcement positif : chaque action est associée à une probabilité d'utilisation dans le futur et le choix d'une action implique un renforcement de sa probabilité qui est proportionnel au paiement reçu. Börgers et Sarin (1997) analysent les propriétés d'une règle inspirée de Bush et Mosteller et proposée par Cross (1973). Au début de chaque période, l'agent observe l'action qu'il a jouée et le paiement qu'il a reçu lors de la période précédente. Ensuite, il révise les probabilités avec lesquelles les actions sont choisies. Afin que le processus soit bien défini, Börgers et Sarin supposent que les paiements appartiennent à l'intervalle ouvert (0, 1). Soit $e_i^t(x_i)$, la fonction de Kronecker de l'action jouée à la période t définie de la manière suivante :

$$e_i^t(x_i) = \begin{cases} 1 \text{ si } x_i \text{ est jouée en } t, \\ \\ 0 \text{ si } x_i \text{ n'est pas jouée en } t. \end{cases}$$

Sachant que le paiement u_i^t est perçu à la période t, l'agent i révise la probabilité avec laquelle il choisit l'action $x_i \in X_i$ lors de la période $t+1$ de la manière suivante :

$$p_i^{t+1}(x_i) = e_i^t(x_i)u_i^t + (1 - u_i^t)p_i^t(x_i). \qquad (1)$$

Pour bien saisir le mécanisme à l'œuvre dans ce modèle, étudions comment la probabilité associée à chaque action varie d'une période à l'autre. Notons $\delta p_i^t(x_i) = p_i^{t+1}(x_i) - p_i^t(x_i)$ la variation de la probabilité associée à l'action x_i entre les périodes t et $t+1$. Par de simples calculs, on obtient

$$\delta p_i^t(x_i) = u_i^t \left(e_i^t(x_i) - p_i^t(x_i)\right).$$

Deux points méritent d'être soulignés. Premièrement, quand une action est jouée, sa probabilité augmente quel que soit le paiement reçu : la variation δp_i^t

(3) Une revue détaillée de cette littérature est proposée par Young (2004).

(x_i) est positive dès lors que x_i est choisie en t. La procédure de renforcement agit donc toujours de manière positive sur l'action choisie. Et, inversement, la probabilité d'une action qui n'est pas jouée décroît systématiquement. Deuxièmement, les variations enregistrées entre la période t et la période t+1 sont proportionnelles au paiement reçu lors de la période t : un paiement élevé conduit à de fortes variations des probabilités.

Börgers et Sarin (1997) étudient cette règle comportementale dans le cadre de jeux finis à deux joueurs. Pour ce faire, ils ont recours notamment à des méthodes d'approximation des processus stochastiques. Précisément, ils établissent que la dynamique du réplicateur (4) approxime le processus de renforcement défini par (1) sur des horizons temporels bornés. En revanche, les propriétés asymptotiques des deux modèles sont différentes. Ainsi, Börgers et Sarin montrent que dans n'importe quel jeu fini à deux joueurs, le processus de renforcement défini par (1) converge avec une probabilité un vers un profil de stratégies pures. Or, ce n'est pas nécessairement le cas du réplicateur. Par exemple, considérons le jeu à deux joueurs et deux stratégies défini par la matrice de paiement suivante :

	0	1
0	(0.4, 0.2)	(0.2, 0.4)
1	(0.2, 0.4)	(0.4, 0.2)

Ce jeu possède un unique équilibre de Nash en stratégies mixtes et n'admet pas d'équilibre de Nash en stratégies pures. La dynamique du réplicateur cycle autour de l'équilibre de Nash en stratégies mixtes. Cet exemple révèle que le processus de renforcement défini par (1) peut converger vers un profil qui n'est pas un équilibre de Nash. La divergence de comportement asymptotique entre le processus de renforcement et la dynamique du réplicateur tient à la constance, à travers le temps, des variations impulsées sur les probabilités par un paiement donné. Précisément, pour deux périodes données t_1 et t_2 avec $t_1 \neq t_2$, si $p_i^{t1}(x_i) = p_i^{t2}(x_i)$ et $u_i^{t1} = u_i^{t2}$, alors on obtient $\delta p_i^{t1}(x_i) = \delta p_i^{t2}(x_i)$. En conséquence, la probabilité que le processus de renforcement dévie significativement de la dynamique du réplicateur est constante dans le temps. À long terme, cet événement se produit avec la probabilité un. Il ressort donc que la vitesse d'ajustement des probabilités tient un rôle crucial. Afin d'améliorer l'approximation du processus stochastique et les résultats de convergence obtenus, le processus de renforcement peut être défini de façon à éliminer la vitesse

(4) La dynamique du réplicateur est issue de la biologie. Elle permet d'étudier l'évolution de la répartition de stratégies au sein d'une ou plusieurs populations d'individus. Pour une présentation détaillée de cette littérature, cf. Weibull (1995).

constante d'ajustement des probabilités au profit d'une vitesse d'ajustement qui décroisse avec le temps.

Arthur (1993) introduit une variante du processus de renforcement de Cross. Il suppose que la détermination de la probabilité de choisir une action donnée à la période t intègre le nombre de périodes de temps écoulées. Formellement, notons $\xi_i^t(x_i)$ la propension de l'agent i à jouer l'action x_i à la période t. Soient une constante ν telle que $0 \leq \nu \leq 1$ et une constante C telle que $C = \sum_{x_i' \in X_i} \xi_i^1(x_i')$. Supposons, d'une part, que toutes les propensions initiales sont positives et, d'autre part, que tous les paiements sont positifs. Sachant que l'agent i a reçu le paiement u_i^t, la propension à choisir l'action x_i en $t + 1$ est définie comme :

$$\xi_i^{t+1}(x_i) = \frac{C(\xi_i^t(x_i) + e_i^t(x_i)u_i^t)(t+1)^\nu}{Ct^\nu + u_i^t}.$$

Cette formulation permet d'assurer que la somme des propensions sur l'ensemble des actions est égale à Ct^ν à n'importe quelle période t. La probabilité pour l'agent i de choisir l'action x_i à la période $t + 1$ est proportionnelle à la propension de cette action. Formellement, la probabilité que l'agent i choisisse l'action x_i en $t + 1$ est donnée par :

$$p_i^{t+1}(x_i) = \xi_i^{t+1}(x_i)/C(t+1)^\nu \tag{2}$$

L'ajustement de la probabilité de l'action x_i entre les périodes t et $t + 1$ est alors :

$$\delta p_i^t(x_i) = \frac{u_i^t}{Ct^\nu + u_i^t}\left(e_i^t(x_i) - p_i^t(x_i)\right).$$

La vitesse d'ajustement des probabilités dépend donc des valeurs des constantes C et ν. Cette variante du modèle de renforcement permet de bien saisir quelle est la sensibilité des résultats à la vitesse d'ajustement des probabilités. Pour $\nu = 0$, la vitesse d'ajustement est constante et de l'ordre de $1/C$. Dans ce cas, le résultat coïncide avec celui obtenu par Börgers et Sarin (1997) : le processus de renforcement défini par (2) peut converger vers un profil de stratégies qui n'est pas un équilibre de Nash. En posant $\nu \neq 0$, la vitesse d'ajustement est décroissante dans le temps et de l'ordre de $1/Ct^\nu$. Cependant, cette décroissance n'est pas toujours suffisante pour empêcher le processus défini par (2) de se bloquer sur un profil qui n'est pas un équilibre de Nash. Ainsi, Posch (1997) étudie le processus de renforcement défini par (2) dans la classe des jeux à deux joueurs et deux stratégies en supposant $1/2 \leq \nu \leq 1$. Il montre que le processus de renforcement est gouverné par la dynamique du réplica-

teur. Il établit que si $\nu < 1$, alors le processus peut converger vers un profil qui n'est pas un équilibre de Nash. Et, ce résultat persiste si $\nu = 1$ et que la constante C est inférieure à tous les paiements du jeu. Dans ces deux cas, il existe toujours une possibilité de blocage sur un profil qui n'est pas un équilibre du fait de la vitesse d'ajustement des probabilités qui est trop élevée lors des premières périodes. Et, par la suite, la décroissance de la vitesse d'ajustement ne permet pas d'échapper au blocage. D'un point de vue plus technique, ces cas renvoient à une situation où le processus de renforcement défini par (2) converge avec une probabilité positive vers un point fixe instable de la dynamique du réplicateur. En revanche, lorsque $\nu = 1$ et que la constante C est supérieure à tous les paiements du jeu, Posch montre que le processus de renforcement défini par (2) converge vers un équilibre de Nash strict, pourvu que ce dernier existe. Autrement dit, avec des restrictions suffisantes imposées sur ν et C, la possibilité de converger vers un point fixe instable de la dynamique du réplicateur est exclue.

D'une manière générale, cet ensemble de résultats suggère deux remarques. Premièrement, la vitesse d'ajustement des probabilités doit être suffisamment faible et décroissante avec le nombre d'interactions. Au niveau de la rationalité individuelle, ce double impératif révèle deux enseignements. D'un côté, les variations des probabilités d'activation de chaque action disponible ne doivent pas avoir une amplitude trop importante. Dans le cas contraire, un effet d'irréversibilité peut survenir : un blocage sur une action sous-efficiente peut intervenir dès les premières périodes du processus. En conséquence, une phase d'exploration de l'éventail des actions disponibles doit pouvoir être réalisée lors des itérations initiales du processus sans renforcement excessif d'une action particulière. D'un autre côté, la décroissance de la vitesse d'ajustement traduit le fait que les probabilités de choix des actions deviennent moins erratiques au fur et à mesure que les situations de décision se succèdent. Ceci signifie que le poids accordé à la dernière itération diminue lorsque l'expérience s'accumule. Tout se passe comme si un effet de mémoire opérait et permettait de relativiser l'importance de la dernière période. Deuxièmement, les restrictions à apporter à la vitesse d'ajustement des probabilités sont liées aux paiements du jeu. Ce point découle directement de la procédure de calcul des propensions à choisir les actions qui incorporent les paiements reçus. La version suivante du processus de renforcement confirme cette intuition en liant directement la vitesse d'ajustement aux paiements reçus cumulés.

Laslier, Topol et Walliser (2001) présentent un modèle où les propensions attachées aux actions reflètent directement les paiements reçus dans le passé. Précisément, sachant que l'agent i a reçu le paiement u_i^t à la période t, la propension de i à jouer x_i en $t + 1$ est :

$$\theta_i^{t+1}(x_i) = \theta_i^t(x_i) + e_i^t(x_i)u_i^t.$$

À la période initiale, la propension initiale à jouer x_i, $\theta_i^1(x_i)$, est supposée positive. Les paiements sont également supposés positifs. La probabilité que i choisisse x_i à la période $t+1$ est déterminée de la manière suivante :

$$p_i^{t+1}(x_i) = \theta_i^{t+1}(x_i) \,/\, \sum_{x_i' \in X_i} \theta_i^{t+1}(x_i'). \tag{3}$$

L'ajustement de la probabilité de l'action x_i entre les périodes t et $t+1$ est alors :

$$\delta p_i^t(x^i) = \frac{u_i^t}{\sum_{x_i' \in X_i} \theta_i^1(x_i') + \sum_{s \le t} u_i^s} \left(e_i^t(x_i) - p_i^t(x_i) \right) \tag{4}$$

La vitesse d'ajustement des probabilités est donc décroissante avec le temps puisque le dénominateur de (4) qui incorpore les paiements cumulés perçus croît avec le temps. Cette vitesse d'ajustement est donc propre à chaque agent et stochastique : elle reflète l'expérience accumulée par chaque agent. Comparativement à la vitesse d'ordre $1/Ct$ obtenue avec le processus de renforcement d'Arthur lorsque $\nu = 1$, on notera que la vitesse d'ajustement définie par (4) est d'ordre inférieur lorsque C est inférieure à tous les paiements du jeu et est d'ordre supérieur lorsque C est supérieure à tous les paiements du jeu. Une telle vitesse d'ajustement est censée favoriser l'approximation du processus de renforcement par des dynamiques déterministes.

Beggs (2005) et Hopkins et Posch (2005) étudient le processus de renforcement défini par (3). Ils montrent que ce processus est proche d'une forme ajustée de la dynamique du réplicateur. Beggs, d'une part, et Hopkins et Posch, d'autre part, établissent que le processus de renforcement défini par (3) ne converge pas vers les points fixes instables de la dynamique du réplicateur ajusté. Pour un sous-ensemble de la classe des jeux à deux joueurs, Hopkins et Posch montrent que le processus de renforcement défini par (3) converge vers un équilibre de Nash. Beggs complète partiellement cette analyse pour la classe des jeux à deux joueurs et deux stratégies : il établit un résultat de convergence similaire pour les jeux à somme nulle avec un unique équilibre de Nash.

III. — RENFORCEMENT ET SEUIL D'ASPIRATION

Les modèles de renforcement présentés précédemment font intervenir les paiements reçus lors des périodes passées. Et, tous les paiements sont supposés positifs. En conséquence, un point commun à tous ces modèles est de faire intervenir un effet de renforcement positif des stratégies jouées. La variation de la probabilité associée à l'action choisie à la période précédente est toujours positive. L'incorporation d'une telle prime à l'utilisation peut conduire à des résultats discutables. Ainsi, dans un jeu contenant une stratégie strictement dominée, le choix de cette stratégie à une période donnée engendre, à la période suivante, une amélioration de sa pondération. De plus, deux stratégies rapportant les mêmes paiements peuvent être plus ou moins valorisées selon les

fréquences d'utilisation plus ou moins élevées : la stratégie la plus fréquemment choisie est associée à la probabilité la plus élevée. En d'autres termes, la démarquation entre les deux stratégies n'est pas absolument fondée sur l'observation des paiements. *A contrario*, certains modèles sont élaborés de façon à prendre en compte des effets de renforcement négatifs.

Börgers et Sarin (2000) proposent un modèle de renforcement qui intègre un seuil d'aspiration. À la différence des modèles précédents, ce ne sont pas les paiements reçus qui sont directement pris en considération mais des paiements relatifs établis en comparaison avec le seuil d'aspiration. Lorsque l'action est jugée non satisfaisante au regard du seuil d'aspiration, alors elle subit un renforcement négatif. À l'instar de Börger et Sarin (1997), les paiements sont restreints à l'intervalle (0, 1). Et les valeurs possibles du seuil d'aspiration sont supposées appartenir à l'intervalle [0, 1]. Deux scénarios alternatifs sont envisagés : soit le seuil d'aspiration est supposé exogène et immuable dans le temps, soit il est considéré comme endogène et est ajusté périodiquement. Formellement, notons a_i^{t+1} le seuil d'aspiration de l'agent i lors de la période $t + 1$. Ce seuil est déterminé de la façon suivante :

$$a_i^{t+1} = \alpha_i u_i^t + (1 - \alpha_i) a_i^t \tag{5}$$

où $\alpha_i \in [0, 1]$ est un paramètre mesurant le pas d'ajustement du seuil d'aspiration. Si le seuil est considéré comme exogène, alors il vient $\alpha_i = 0$ pour chaque agent $i \in N$. Si le seuil est endogène, alors α_i est positif et, à chaque période, le seuil d'aspiration est ajusté en fonction du dernier paiement reçu par l'agent. Ainsi, à la période $t + 1$, le seuil décroît lorsque le paiement reçu à la période t, u_i^t, est inférieur à la valeur précédente du seuil, a_i^t, et inversement. En d'autres termes, le seuil d'aspiration de l'agent i reflète son expérience en matière de paiements reçus.

Sachant que le paiement u_i^t est perçu par l'agent i à la période t, l'ajustement des probabilités suit deux procédures distinctes selon que u_i^t est jugé satisfaisant ou non. Dans le premier cas, c'est-à-dire si $u_i^t \geq a_i^t$, la probabilité de choisir l'action x_i à la période $t + 1$ est déterminée par :

$$p_i^{t+1}(x_i) = e_i^t(x_i)(u_i^t - a_i^t) + \left(1 - (u_i^t - a_i^t)\right) p_i^t(x_i) \tag{6}$$

Comparativement à l'équation (1) définissant le processus de renforcement étudié par Börgers et Sarin (1997), les probabilités associées aux actions dépendent de l'écart constaté entre le paiement reçu et le seuil d'aspiration. L'ajustement de la probabilité de l'action x_i entre les périodes t et $t + 1$ est alors :

$$\delta p_i^t(x_i) = (u_i^t - a_i^t)(e_i^t(x_i) - p_i^t(x_i)).$$

Dans ce cas, les variations de la probabilité associée à l'action jouée à la période précédente sont toujours positives : une action dont le paiement a été évalué comme satisfaisant est systématiquement renforcée de manière positi-

ve. Dans le cas contraire, c'est-à-dire si $u_i^t < a_i^t$, la probabilité de choisir l'action x_i à la période $t + 1$ est déterminée de la façon suivante :

$$p_i^{t+1}(x_i) = (1 - e_i^t(x_i))\,|u_i^t - a_i^t| + \left(1 - |u_i^t - a_i^t|\right) p_i^t(x_i) \qquad (7)$$

L'ajustement de la probabilité de l'action x_i entre les périodes t et $t + 1$ est alors :

$$\delta p_i^t(x_i) = |u_i^t - a_i^t|(1 - e_i^t(x_i) - p_i^t(x_i)).$$

La variation de la probabilité de l'action ayant rapporté un paiement jugé non satisfaisant est donc négative. Un effet de renforcement négatif se met en place au profit des actions qui n'ont pas été choisies à la période précédente. L'agent constatant que le paiement perçu est non satisfaisant se trouve donc moins enclin à utiliser cette action.

Börgers et Sarin (2000) étudient le processus de renforcement défini par (6) et (7) dans le cadre de jeux contre la Nature. Börgers et Sarin traitent de jeux dans lesquels l'agent dispose de deux stratégies pures. Il est supposé, en outre, que chaque stratégie reçoit une pondération initiale positive. De son côté, Nature est caractérisée par une loterie sur un ensemble fini d'états. Les résultats sont à distinguer selon le scénario retenu pour le seuil d'aspiration. Un premier résultat concerne le processus de renforcement défini avec un seuil d'aspiration exogène, *i.e.* avec $\alpha_i = 0$. Börgers et Sarin (2000) montrent que les résultats de convergence obtenus dépendent de manière cruciale de la valeur du seuil d'aspiration. Si le seuil d'aspiration n'est pas trop élevé, alors le processus permet de sélectionner avec une probabilité un l'action qui procure le paiement espéré le plus élevé. Précisément, le processus de renforcement défini par (6) et (7) converge avec la probabilité un vers la stratégie associée au paiement espéré maximal pourvu que le seuil d'aspiration soit inférieur à tous les paiements procurés par cette stratégie. Dans cette situation, seuls des effets de renforcement positifs peuvent se produire en faveur de cette stratégie. Le résultat dérive d'une approximation du processus de renforcement par la dynamique du réplicateur construit non pas sur la base des paiements des actions mais sur la base des bénéfices nets (positifs) relativement au seuil d'aspiration. L'action associée au bénéfice net espéré le plus élevé est alors sélectionnée. Dans le cas contraire où certains paiements sont inférieurs au seuil d'aspiration, le choix de la meilleure stratégie en terme de paiement espéré peut conduire à des paiements jugés insatisfaisants. La stratégie alternative peut alors être choisie avec une probabilité positive.

L'obtention de résultats de convergence pour le processus de renforcement défini avec un seuil d'aspiration endogène nécessite de restreindre la classe des jeux contre la Nature retenue. Börgers et Sarin (2000) considèrent des jeux contre la Nature tels que l'action maximale en terme de paiement espéré possède deux propriétés supplémentaires. Premièrement, l'action est associée à un paiement constant : elle rapporte le même paiement quel que soit l'état de Nature réalisé. Deuxièmement, l'action est dominante : pour n'importe quel

état de Nature, le paiement procuré par cette action est supérieur à celui rapporté par l'action alternative. À titre d'exemple, considérons un jeu contre la Nature dans lequel l'agent a deux stratégies pures à sa disposition, {0, 1}. L'ensemble des états de Nature est $\{\omega_1, \omega_2, \omega_3\}$. Soit la matrice de paiement suivante :

	ω_1	ω_2	ω_3
0	5	5	5
1	1	2	4

La stratégie 0 est dominante et a un paiement constant. Börgers et Sarin (2000) montrent que le processus de convergence défini par (6) et (7) converge avec la probabilité un vers l'action maximale en terme de paiement espéré si cette action a un paiement constant et est dominante. En outre, Börgers et Sarin établissent que ces deux propriétés sont nécessaires pour l'obtention de ce résultat. Un argument essentiel pour ce résultat consiste dans le fait qu'il existe des valeurs du seuil d'aspiration telles que l'action maximale est systématiquement renforcée positivement quand elle est choisie tandis que l'action alternative peut être renforcée négativement lorsqu'elle est jouée. Ces valeurs renvoient à des seuils d'aspiration qui se situent dans la zone des paiements compris entre le paiement constant de l'action maximale et le paiement minimum de l'action alternative. De plus, par les propriétés de l'action maximale, la situation dans laquelle l'action maximale est renforcée négativement quand elle est choisie alors que l'action alternative pourrait être renforcée positivement si elle était jouée, ne peut pas survenir. Précisément, pour tout état de Nature, il est impossible que l'action maximale soit renforcée négativement du fait d'un paiement non satisfaisant (*i.e.* quand elle est jouée) dès lors que le seuil d'aspiration est inférieur au paiement espéré de cette action. Et, il est impossible que le seuil d'aspiration converge vers une valeur supérieure à ce paiement.

Au total, la prise en considération d'un seuil d'aspiration exogène ou endogène complexifie sensiblement l'analyse du processus de renforcement. En conséquence, Börgers et Sarin (2000) introduisent un certain nombre de restrictions. Leurs résultats reposent notamment sur certaines conditions assurant qu'aucun effet de renforcement négatif n'opère sur l'action optimale. Dans le cas où le seuil d'aspiration est exogène, une telle condition est exprimée par une restriction portant sur la valeur du seuil. Lorsque le seuil est endogène, il est nécessaire de doter l'action optimale de propriétés particulières. Seules certaines catégories de jeux contre la Nature peuvent donc être étudiées. L'obtention de résultats pour une classe plus large de situations nécessiterait alors une modification du critère de convergence retenu. En revanche, notons que le résultat obtenu par Börgers et Sarin (2000) avec un seuil d'aspiration endogène ne dépend pas de la valeur du paramètre α_i. En d'autres termes, le résultat ne dépend pas de l'importance que l'agent attribue au paiement de la

dernière itération relativement à l'expérience passée reflétée dans la valeur précédente du seuil d'aspiration.

Karandikar *et alii* (1998) considèrent une règle comportementale présentant de nombreuses similitudes avec celle étudiée par Börgers et Sarin (2000). En revanche, l'analyse est menée dans un cadre d'interaction stratégique entre agents. Les choix de stratégies de chaque agent sont fondés sur des comparaisons entre les paiements reçus et un seuil d'aspiration. Karandikar *et alii* (1998) considèrent un seuil d'aspiration endogène qui évolue de période en période selon la procédure décrite par (5) où $\alpha_i \in (0, 1)$ pour chaque agent i. Toutefois, comparativement à Börgers et Sarin (2000), Karandikar *et alii* (1998) envisagent un scénario supplémentaire dans lequel l'ajustement des seuils d'aspiration fait l'objet d'erreur. Précisément, à chaque période, les agents ajustent leur seuil d'aspiration en respectant la règle décrite dans (5) avec une probabilité $1 - \epsilon$. Et, avec la probabilité complémentaire, ils sont sujets à des tremblements. Le seuil d'aspiration prend alors une valeur déterminée aléatoirement au sein d'un intervalle contenant l'ensemble des paiements possibles.

La détermination des probabilités de choix des actions résulte largement de la confrontation entre la valeur courante du seuil d'aspiration et le paiement enregistré lors de la dernière période. Deux situations sont donc à considérer selon que le paiement est jugé satisfaisant ou non au regard du seuil d'aspiration. Supposons tout d'abord que le dernier paiement perçu est satisfaisant. Sachant que le paiement reçu par l'agent i à la période t est u_i^t tel que $u_i^t \geq a_i^t$, la probabilité que i choisisse l'action x_i en $t + 1$ est :

$$p_i^{t+1}(x_i) = e_i^t(x_i). \tag{8}$$

Ceci signifie qu'une action ayant généré un paiement satisfaisant est jouée avec la probabilité un. Autrement dit, un agent reconduit systématiquement l'action dont le paiement a été satisfaisant. À présent, considérons la situation où le dernier paiement reçu a été inférieur au seuil d'aspiration. L'agent insatisfait a alors une probabilité strictement positive d'opter en faveur d'une stratégie alternative. Et, cette probabilité dépend de l'écart constaté entre le paiement reçu et le seuil d'aspiration. Karandikar *et alii* (1998) concentrent leur attention sur un jeu dans lequel les agents disposent seulement de deux actions. Afin de définir formellement la probabilité de chaque action, introduisons une fonction $f : \mathbb{R} \rightarrow (0, 1)$ qui est continue et décroissante. Sachant que le paiement reçu par l'agent i à la période t est u_i^t tel que $u_i^t < a_i^t$, la probabilité que i choisisse l'action x_i en $t + 1$ est :

$$p_i^{t+1}(x_i) = e_i^t(x_i) f(a_i^t - u_i^t) + (1 - e_i^t(x_i))(1 - f(a_i^t - u_i^t)). \tag{9}$$

Plus l'écart entre le seuil d'aspiration et le dernier paiement est élevé, plus la probabilité de choisir l'action jouée à la période précédente est faible ; et, inversement, plus la probabilité de jouer l'action alternative est élevée. Notons que Karandikar *et alii* (1998) supposent que pour tout écart entre le seuil d'as-

piration et le paiement reçu, la probabilité de conserver l'action précédente est strictement positive. Ceci revient à supposer qu'il existe une certaine inertie dans le basculement d'une action vers l'autre. À la différence des modèles présentés précédemment, observons que la probabilité de choisir une action à une période donnée ne dépend pas de la probabilité attachée à cette action lors de la période précédente. Ceci signifie que l'expérience accumulée dans le passé en matière de paiements n'entre pas directement dans la détermination de la probabilité de choisir cette action. En revanche, les paiements passés sont intégrés dans l'analyse par l'intermédiaire du seuil d'aspiration.

Karandikar *et alii* (1998) étudient le processus défini par (8) et (9) en référence à la classe des jeux à deux joueurs et deux stratégies. En particulier, ils retiennent deux types de jeux. En premier lieu, ils considèrent un dilemme du prisonnier. Un exemple d'un tel jeu est fourni par la matrice de paiement suivante :

	0	1
0	(2, 2)	(0, 3)
1	(3, 0)	(1, 1)

Le profil de stratégies pures (1, 1) est l'unique équilibre en stratégies dominantes de ce jeu. En second lieu, Karandikar *et alii* s'intéressent à un jeu d'intérêt commun avec deux équilibres de Nash en stratégies pures. À titre d'exemple, prenons le jeu dont la matrice de paiement est

	0	1
0	(3, 3)	(0, 1)
1	(1, 0)	(2, 2)

Les profils de stratégies pures (0, 0) et (1, 1) sont les équilibres de Nash stricts de ce jeu. L'équilibre (0, 0) domine au sens de Pareto.

Deux types de résultats sont obtenus selon le scénario retenu quant à la présence ou l'absence de tremblement au niveau de la fixation du seuil d'aspiration. Un premier résultat se penche sur la convergence du processus lorsque les tremblements sont absents. Lorsque $\epsilon = 0$, Karandikar *et alii* (1998) montrent que, dans chaque type de jeu, le processus défini par (8) et (9) converge vers n'importe quel profil de stratégies pures tel que le seuil d'aspiration est précisément égal au paiement associé au profil considéré. Soulignons qu'à la différence de Börgers et Sarin (2000), il est possible que le processus se bloque sur un profil contenant au moins une action qui est strictement dominée (par exemple dans le dilemme du prisonnier). En effet, si la valeur initiale du seuil d'aspiration est suffisamment basse, à savoir inférieure au paiement reçu lorsque l'action considérée est jouée, alors chaque joueur reconduit systématiquement sa stratégie et le seuil d'aspiration converge vers le paiement.

L'introduction de tremblements dans la procédure d'ajustement des seuils d'aspiration permet d'éviter de tels blocages. Karandikar *et alii* (1998) étudient le comportement asymptotique du processus défini par (8) et (9) quand $\epsilon \to 0$. Ils établissent que le profil de stratégies pures (0, 0) est joué avec une probabilité proche de un pourvu que la vitesse d'ajustement du seuil d'aspiration pour chaque joueur *i* soit suffisamment lente (*i.e.* $\alpha_i \to 0$) (5). Dans le cadre d'un dilemme du prisonnier, ce résultat signifie que le processus ne converge pas vers un équilibre de Nash mais que les joueurs choisissent le profil qui maximise leurs paiements. Pour le jeu d'intérêt commun, le processus converge vers l'équilibre de Nash strict qui domine au sens de Pareto.

À l'issue de cette section, deux avantages résultant de l'introduction d'un seuil d'aspiration (exogène ou endogène) peuvent être mis en évidence. Premièrement, dans le modèle de Börgers et Sarin (2000), le seuil d'aspiration en autorisant des effets de renforcement négatifs permet d'affiner l'évaluation de chacune des actions disponibles. Deuxièmement, dans le modèle de Karandikar *et alii* (1998), le seuil d'aspiration endogène rend non systématique la révision de l'action jouée : un changement d'action est envisagé seulement lorsque le paiement reçu lors de la dernière itération est inférieur au seuil d'aspiration. De ce point de vue, l'incorporation du seuil d'aspiration dans la procédure de décision satisfait un objectif énoncé par H. Simon : le seuil d'aspiration permet de réaliser une économie dans l'utilisation des ressources cognitives. Lorsque l'action courante s'avère satisfaisante, la procédure de décision se résume, d'une part, à une comparaison entre le seuil d'aspiration et le dernier paiement reçu et, d'autre part, à un ajustement de la valeur du seuil d'aspiration.

IV. — MODÈLES D'IMITATION

Dans cette dernière section, nous tournons notre attention vers des règles comportementales dans lesquelles les agents exploitent des informations factuelles rétrospectives concernant d'autres agents. Sur la base de ces informations, les agents adoptent des comportements mimétiques. Autrement dit, ils sont susceptibles d'adopter une action qui a été jouée. *A contrario*, ils sont dans l'incapacité de choisir une action qui n'appartient pas à leur champ d'observation. Des informations de nature et en quantité différentes peuvent être

(5) Notons que le critère de convergence appliqué ici est plus fort que dans le cas du critère de stabilité stochastique présenté dans l'article suivant de Baron et Solal puisque Karandikar *et alii* requièrent que les pondérations figurant dans la distribution stationnaire limite soient quasiment concentrées sur un état unique. L'application du critère de stabilité stochastique conduit simplement à rechercher les états recevant un poids positif dans la distribution stationnaire limite. Plusieurs états peuvent alors être stochastiquement stables.

incorporées dans les procédures de décisions individuelles. Naturellement, la base informationnelle minimale pour le développement de comportements d'imitation est la connaissance des actions choisies par au moins un autre agent lors d'une période antérieure. Nous présentons uniquement des modèles intégrant des informations portant sur les paiements reçus par d'autres agents. Cette restriction implique en particulier que nous laissons de côté certains modèles qui fondent les décisions individuelles sur des critères tels que la sélection aléatoire ou la popularité d'une action. Cependant, il subsiste un fort degré de liberté dans la spécification exacte de la règle adoptée.

Considérons la situation d'une population N d'agents qui jouent chacun le même jeu contre la Nature, caractérisée par une loterie sur un ensemble fini d'états Ω. Autrement dit, tous les agents disposent du même ensemble d'actions et de la même fonction de paiement. Supposons que les agents ont accès à l'ensemble des actions jouées et des paiements reçus dans la population lors de la période précédente. Une règle qui est susceptible d'être appliquée dans ce contexte informationnel est la règle d'imitation de l'action associée au meilleur paiement reçu. Soit S^t l'ensemble des actions ayant procuré le meilleur paiement dans la population à la période t :

$$S^t = \{x_j^t, j \in N : u_j(x_j^t, \omega_j^t) = \max_{k \in N} u_k(x_k^t, \omega_k^t)\}$$

où $\omega_j \in \Omega$ désigne l'état dans lequel l'agent j se situe à la période t et qui résulte de la loterie qui caractérise Nature. L'indice j dans ω_j signifie simplement que Nature effectue un tirage pour chaque agent j de sorte que deux agents peuvent probablement évoluer dans deux états de Nature différents au même instant. L'ensemble S^t n'est pas indicé par une variable individuelle puisque les observations sur les stratégies jouées et les paiements sont communes à tous les individus. En d'autres termes, les informations concernant les actions jouées et les paiements reçus sont publiques. Étant donné S^t, la probabilité que l'agent i choisisse l'action x_i à la période $t + 1$ lorsqu'il agit en conformité avec la règle d'imitation de la stratégie attachée au meilleur paiement est donnée par :

$$p_i^{t+1}(x_i) = \begin{cases} 1/|S^t| & \text{si } x_i \in S^t, \\ 0 & \text{autrement} \end{cases}$$

Un exemple simple permet de constater que cette règle d'imitation peut conduire à des choix inefficients. Considérons le jeu contre la Nature à deux stratégies et deux états introduit par Schlag (1998a) et défini par la matrice de paiement suivante :

	ω_1	ω_2
0	4.5	4.5
1	5	2.5

Supposons que la loterie caractérisant Nature accorde la probabilité 2/3 à l'état ω_1 et la probabilité complémentaire à l'état ω_2. Ceci implique que l'action 0 rapporte un paiement espéré plus élevé que l'action 1. Cependant, supposons qu'à une période un agent choisisse l'action 1. Avec la probabilité 2/3, cet agent reçoit le paiement le plus élevé possible. Si tel est le cas, lors de la période suivante, tous les agents imitent l'action 1. Au total, l'action associée au paiement espéré le moins élevé, et donc le moins efficace sur le long terme, est adoptée par la population.

Partant de la mise en évidence de cette difficulté, deux tentatives de résolution peuvent être envisagées. Soit l'information factuelle rétrospective mobilisée est enrichie. En particulier, il est possible d'introduire des effets de mémoire limitée : chaque agent a la capacité de se souvenir des profils de stratégies joués et des paiements reçus lors de plusieurs périodes les plus récentes. Différents critères d'évaluation et donc de sélection des actions sont alors envisageables. En particulier, chaque agent peut opter en faveur de l'action ayant généré le meilleur paiement moyen à travers la population sur l'ensemble des périodes en mémoire. Lorsque la taille de la mémoire est suffisamment importante, les paiements moyens observés tendent à converger vers les paiements espérés. La probabilité de choisir l'action associée au paiement espéré le plus élevé se trouve donc améliorée. Des variantes similaires du modèle d'imitation sont notamment étudiées par Alós-Ferrer (2003) et Josephson et Matros (2004) (6). Soit, au contraire, il est supposé que les agents ont accès à une information moins riche. Par exemple, chaque agent ne peut observer qu'une fraction des stratégies et des paiements reçus dans la population lors de la dernière période. En revanche, cette restriction informationnelle est compensée par la mise en œuvre d'une procédure de choix des actions plus sophistiquée.

Schlag (1998b, 1999) étudie les déclinaisons possibles de la règle d'imitation dans le cadre de jeux contre la Nature et avec un scénario informationnel limité à un ou deux agents observés. Précisément, lorsque le volume d'information est restreint à un agent, à chaque période, chaque agent tire un individu au hasard dans la population et observe la stratégie jouée et le paiement reçu lors de la dernière période. L'agent décide alors d'adopter la stratégie qu'il observe ou bien de conserver sa stratégie courante sur la base d'une comparaison entre son propre paiement et celui obtenu par l'autre agent. En particulier, Schlag met en avant une règle d'imitation proportionnelle qui stipule qu'un agent n'imite pas une stratégie observée ayant rapporté un paiement inférieur à son propre paiement. De plus, en cas d'imitation, la probabilité

(6) Cf. les présentations ultérieures de Baron et Solal et de Billand et Bravard.

d'adoption de la stratégie observée dépend de l'écart de paiement constaté entre les deux stratégies. Formellement, considérons un agent i qui joue l'action x_i et perçoit le paiement u_i^t à la période t. Supposons que les paiements de tous les agents i appartiennent à l'intervalle borné $[\beta, \gamma]$. Postulons que l'agent i a connaissance de cet intervalle. Sachant que l'agent i tire aléatoirement l'agent j au début de la période $t+1$, la probabilité que l'agent i imite l'agent j est :

$$p_i^{t+1}(x_j^t) = \begin{cases} \dfrac{u_j^t - u_i^t}{\gamma - \beta} & \text{si } u_j^t > u_i^t \\ 0 & \text{si } u_j^t \leq u_i^t \text{ et } x_j^t \neq x_i^t \end{cases}$$

Afin de spécifier complètement la règle d'imitation proportionnelle, notons que $p_i^{t+1}(x_j^t) = 1$ si $x_j^t = x_i^t$. Et, la probabilité de choisir une action qui n'est pas observée, c'est-à-dire qui n'est jouée ni par i ni par j à la période t, est nulle.

L'application de la règle d'imitation proportionnelle à l'exemple précédent permet de favoriser l'obtention d'un résultat efficace. Considérons un agent i qui choisit l'action 0 et tire l'agent j qui joue l'action 1. La probabilité avec laquelle i imite j est $(1/2.5) \times (5 - 4.5) \times (2/3) = 2/15$. À l'inverse, si j tire i, alors la probabilité que j imite i est $(1/2.5) \times (4.5 - 2.5) \times (1/3) = 4/15$. Au total, étant donnés les tirages entre agents, la probabilité qu'un agent bascule de l'action 1 à l'action 0 est supérieure à la probabilité d'un mouvement inverse. En conséquence, la règle d'imitation proportionnelle permet de réduire substantiellement la probabilité d'imitation de l'action 1. De ce point de vue, la restriction apportée au niveau informationnel et la mise en place d'une procédure d'imitation plus sophistiquée sont bénéfiques : elles favorisent l'action associée au paiement espéré maximal.

D'une manière plus générale, Schlag propose d'introduire une procédure d'évaluation des règles d'imitation conditionnellement à l'information disponible. Ainsi, Schlag (1998b) se penche sur la question de la sélection des règles satisfaisantes relativement à un critère donné quand les individus ont accès à l'information portant sur un autre agent. Il considère une population d'agents qui jouent tous le même jeu contre la Nature. En conséquence, on a $X_i = X$ pour tous les agents $i \in N$. Et le paiement espéré de chaque action disponible pour tous les agents est immuable dans le temps. Notons $\bar{u}(x_i)$ le paiement espéré de l'action x_i. À chaque période t, il est possible de calculer le paiement espéré moyen de la population. Notons $n^t(x_i)$ le nombre d'agents de la population ayant choisi l'action x_i lors de la période t. Le paiement espéré moyen de la période t est

$$\bar{u}^t = \sum_{x_i \in X} \frac{n^t(x_i)}{n} \bar{u}(x_i).$$

Schlag considère une procédure de tirage aléatoire entre les agents qui est mutuelle : les agents i et j sont tirés par paire et chacun observe la stratégie jouée et le paiement reçu par l'autre agent lors de la période précédente.

Schlag (1998b) propose un critère d'évaluation des règles de décision basé sur les résultats générés par la règle stipulant de conserver son action courante quelles que soient les observations enregistrées. Une spécificité de cette règle est qu'elle engendre l'invariance du paiement espéré moyen entre deux périodes quand elle est employée par tous les agents de la population. En effet, si tous les agents ont recours à cette règle, la proportion d'agents choisissant chacune des stratégies est inchangée et donc le paiement espéré moyen est constant. Une règle est déclarée améliorante (et est donc jugée satisfaisante) si elle procure, lorsqu'elle est adoptée par toute la population, un paiement espéré moyen supérieur ou égal à celui procuré par la règle préconisant de ne jamais changer d'action. Autrement dit, une règle est améliorante si elle permet d'améliorer (faiblement) le paiement espéré moyen d'une période sur l'autre lorsque tous les agents se conforment à cette règle.

Schlag (1998b) opère une sélection parmi les règles comportementales mobilisant des informations concernant un autre agent en appliquant un raisonnement en deux temps. Premièrement, il introduit une propriété caractérisant les règles améliorantes. Afin de présenter cette propriété, introduisons quelques éléments de notation supplémentaires. Sachant que la paire d'agents i, j a été tirée, sachant que le joueur i a joué l'action x_i en t et perçu le paiement u_i^t à la période t et sachant que dans le même temps le joueur j a joué x_j et perçu le paiement u_j^t, la probabilité que le joueur i imite l'agent j et choisisse l'action x_j à la période $t+1$ est notée $p_i^{t+1}(x_j \,|x_i, u_i^t, u_j^t)$. Notons $q(u_i(x_i))$ la probabilité que l'action x_i procure le paiement $u_i(x_i)$. La probabilité globale que l'agent i imite l'agent j sachant que la paire (i, j) a été constituée est

$$p_i^{t+1}(x_j|x_i) = \sum_{u_i(x_i),u_j(x_j)} q(u_i(x_i))q(u_j(x_j)p_i^{t+1}(x_j|x_i, u_i^t, u_j^t).$$

Schlag (1998b) montre que les règles améliorantes sont des règles mimétiques caractérisées, lorsque la paire (i, j) est formée, par la propriété suivante :

$$\left(p_i^{t+1}(x_j|x_i) - p_j^{t+1}(x_i|x_j)\right)(\overline{u(x_j)} - \overline{u(x_i)}) \geq 0. \qquad (10)$$

Une règle améliorante prescrit donc toujours à un individu de choisir entre son action courante et l'action observée ; la probabilité d'opter pour une autre action est nulle. De plus, d'après (10), l'agent dont l'action est associée au paiement espéré le moins élevé est donc le plus enclin à imiter. Deuxièmement, Schlag (1998b) identifie une classe de règles améliorantes, c'est-à-dire de règles d'imitation vérifiant la propriété (10). À la différence de la propriété (10), la caractérisation de cette classe de règles est basée sur des informations disponibles pour les agents, à savoir sur les paiements reçus (et non pas sur les paiements espérés). Si les agents i et j sont appariés à la pério-

de $t + 1$, Schlag établit qu'une règle est améliorante si et seulement si elle vérifie :

$$p_i^{t+1}(x_j|x_i, u_i^t, u_j^t) - p_j^{t+1}(x_i|x_j, u_j^t, u_i^t) = \sigma_{ij}(u_j^t - u_i^t)$$

où $\sigma_{ij} \in [0, 1 / (\gamma - \beta)]$ est le taux de changement d'une action vers l'autre. Si $\sigma_{ij} = 0$, alors la probabilité d'imiter l'autre agent est égale pour les deux agents et indépendante de l'écart de paiements reçus. Si $\sigma_{ij} > 0$, alors l'individu ayant perçu le paiement le moins élevé est le plus enclin à imiter. En plus, la tendance nette à basculer vers la stratégie observée dépend de l'écart de paiements constaté : plus l'action observée rapporte un paiement élevé relativement au paiement reçu, plus la probabilité d'imiter cette action est élevée relativement à la probabilité d'opérer le changement inverse.

Parmi l'ensemble des règles améliorantes, Schlag (1998b) étudie les propriétés d'une règle particulière définie par la valeur maximale du taux de changement, *i.e.* $\sigma_{ij} = 1 / (\gamma - \beta)$ pour n'importe quels agents $i, j \in N$. Il montre que cette règle procure les améliorations du paiement espéré moyen les plus élevées parmi l'ensemble des règles améliorantes. En effet, il apparaît que, lorsqu'une règle améliorante est suivie, l'amélioration du paiement espéré moyen dépend linéairement des taux de changement $(\sigma_{ij})_{i,j \in N}$. En conséquence, la règle élaborée à partir de la valeur $\sigma_{ij} = 0$ pour n'importe quels agents $i, j \in N$ engendre une amélioration attendue égale à 0. Lorsque $\sigma_{ij} > 0$ pour n'importe quels agents $i, j \in N$, alors les règles sont strictement améliorantes. Et la règle dominante en termes d'amélioration correspond à la valeur $\sigma_{ij} = 1 / (\gamma - \beta)$ pour n'importe quels agents $i, j \in N$. En outre, Schlag (1998b) prouve que la règle améliorante dominante possède la propriété de ne jamais prescrire à un agent d'imiter une action dont le paiement observé est inférieur au paiement qu'il a reçu. Enfin, Schlag (1998b) établit que la règle améliorante dominante conduit les agents d'une population de grande taille à sélectionner les actions de paiement espéré maximal parmi les actions choisies lors de la période initiale. La dépendance vis-à-vis des conditions initiales résulte de l'impossibilité de choisir une action qui n'est pas jouée. Au total, si toutes les actions sont jouées initialement, alors les actions de paiement espéré maximal seront jouées avec une forte probabilité. Pour établir ce résultat, Schlag approxime, sur un horizon temporel borné et pour une population de grande taille, le processus défini par la règle améliorante dominante par la dynamique du réplicateur.

Il ressort de cette analyse que l'obtention des meilleurs résultats en termes de paiements passe par l'intégration de procédures de décision « sophistiquées » dans lesquelles le choix d'une action requiert un traitement des informations enregistrées sur les paiements reçus. Cette intuition est confirmée par les résultats de Schlag (1999) qui aborde une problématique similaire à celle développée dans Schlag (1998b) pour le cas où l'échantillon observé contient deux individus. Les procédures à envisager afin de définir une règle améliorante nécessitent la mise en place de plusieurs comparaisons par paires entre les actions et leurs paiements respectifs. Face à cette tendance à la sophisti-

quation des règles, il est alors légitime de s'interroger sur les motifs susceptibles de conduire les agents à adopter de telles règles.

V. – CONCLUSION

Cette contribution présente certaines voies de modélisation des décisions individuelles caractérisées par une rationalité limitée qui s'exprime au travers de l'hypothèse de comportements adaptatifs. Elle ne constitue, cependant, qu'une vision partielle de l'ensemble des procédures de décision étudiées dans la littérature récente en théorie des jeux. De plus, comme le suggère la présentation précédente des procédures de renforcement et d'imitation, de multiples variantes de chaque modèle sont envisagées. D'un certain point de vue, cette caractéristique met en évidence la difficulté à formaliser des procédures de décision (basées sur des hypothèses restrictives en termes de connaissances et de capacités cognitives) produisant des résultats robustes, c'est-à-dire assurant la convergence dans de nombreuses classes de jeux vers l'ensemble des équilibres de Nash (7). D'un autre côté, une telle multiplicité révèle la richesse et l'important potentiel de cette approche. Ainsi, l'éventail des modèles de comportements adaptatifs permet de représenter des procédures de décision dans des contextes économiques variés. Notamment, les modèles faisant appel à des règles mimétiques s'avèrent particulièrement adaptés pour l'étude des comportements des firmes et des phénomènes qui en découlent en économie industrielle (8).

(7) Pour une discussion plus détaillée de ce point, cf. Young (2004).

(8) Sur ce point, se référer par exemple à l'article ci-après de Billand et Bravard.

Voir bibliographie page suivante

BIBLIOGRAPHIE

ALÓS-FERRER C. (2003), « Learning, memory, and inertia », mimeo.

ARTHUR W.-B. (1993), « On designing economic agent that behaves like human agents », *Journal of Evolutionary Economics*, 3, 1-22.

BEGGS A.-W. (2005), « On the convergence of reinforcement learning », *Journal of Economic Theory*, 122, 1-36.

BÖRGERS T., SARIN R. (1997), « Learning through reinforcement and replicator dynamics », *Journal of Economic Theory*, 77, 1-14.

BÖRGERS T., SARIN R. (2000), « Naive reinforcement learning with endogenous aspirations », *International Economic Review*, 41, 921-950.

BUSH R., MOSTELLER F. (1955), « Stochastic models for learning ». Wiley and Sons, New York.

CROSS J. (1973), « A stochastic learning model of economic behavior », *Quarterly Journal of Economics*, 87, 239-266.

ESTES W.K. (1950), « Toward a statistical theory of learning », *Psychological Review*, 57, 94-107.

HOPKINS E., POSCH M. (2005), « Attainability of boundary points under reinforcement learning », *Games and Economic Behavior*, 53, 110-125.

JOSEPHSON J., MATROS A. (2004), « Stochastic imitation in finite games », *Games and Economic Behavior*, 49, 244-259.

KARANDIKAR R., MOOKHERJEE D., RAY D., VEGA-REDONDO F. (1998), « Evolving aspirations and cooperation », *Journal of Economic Theory*, 80, 292-331.

LASLIER J.-F., TOPOL R., WALLISER B. (2001), « A behavioral learning process in games », *Games and Economic Behavior*, 37, 340-366.

POSCH M. (1997), « Cycling in a stochastic learning algorithm for normal form games », *Journal of Evolutionary Economics*, 7, 193-207.

SCHLAG K. (1998a), « Justifying imitation », mimeo.

SCHLAG K. (1998b), « Why imitate, and if so, how ? », *Journal of Economic Theory*, 78, 130-156.

SCHLAG K. (1999), « Which one should I imitate ? », *Journal of Mathematical Economics*, 31, 493-522.

THORNDIKE E.-L. (1898), « Animal intelligence : an experimental study of the associative processes in animals », *Psychological Review*, 8.

WEIBULL J. (1995), « Evolutionary game theory ». Cambridge : MIT Press.

YOUNG H.-P. (2004), « Strategic learning and its limits ». Oxford : Oxford University Press.