
L'émergence de comportements de contrôle chez des agents sélectionnistes leur permet de résoudre le dilemme du prisonnier.

Jérémie Jozefowicz

*Unité de Recherche sur l'Evolution du comportement et de l'Apprentissage, Université de Lille3, BP 149
59653 Villeneuve d'Ascq cedex*

Jean-Claude Darcheville

*Unité de Recherche sur l'Evolution du comportement et de l'Apprentissage, Université de Lille3, BP 149
59653 Villeneuve d'Ascq cedex*

Philippe Preux

Laboratoire d'Informatique du Littoral, Université du Littoral, BP 719, 62228 Calais cedex

RESUME. *Nous proposons une approche sélectionniste du dilemme du prisonnier dans laquelle des agents adaptatifs simples, dérivés de recherches empiriques en psychologie de l'apprentissage, n'utilisant aucune règle mais dont les comportements sont sélectionnés en fonction de leurs conséquences apprennent à coopérer. Cela est possible car le répertoire de chaque agent ne se limite pas aux comportements de coopération et de non-coopération mais inclut également des comportements de contrôle de l'autre. En supposant que les mécanismes de renforcement des comportements par leurs conséquences sont suffisants pour permettre le maintien de la coopération dans un dilemme du prisonnier à condition que le répertoire comportemental des agents soient étendus, ce travail se distingue d'autres recherches, telles que celles d'Axelrod (1992), qui étudiaient la manière dont la sélection naturelle avaient pu favoriser la propagation de stratégie dans une population.*

MOTS-CLES : *dilemme du prisonnier, sélection des comportements par leurs conséquences, contingences de renforcement, coopération, comportements de contrôle.*

Introduction.

Le sélectionnisme est une approche de la complexité qui considère celle-ci comme le produit cumulatif de processus de variation qui génèrent au sein d'une population tout un ensemble de variantes et de processus de sélection qui sélectionnent au sein de cette population certaines de ces variantes. Proposées pour la première fois par Darwin pour expliquer l'émergence de la complexité et de la diversité dans les règnes animal et végétal au cours de l'évolution, des approches sélectionnistes ont depuis été proposées en sciences cognitives afin de comprendre l'acquisition des comportements et le fonctionnement cérébral [DONAHOE, BURGOS ET PALMER 93] [EDELMAN 92] [SKINNER 81]. Ainsi, en psychologie, selon Skinner, les raisons pour lesquelles un organisme émet un comportement donné s'expliquent par les relations qui existent entre l'environnement avant l'émission du comportement, le comportement lui-même et les conséquences de ce comportement. Skinner appelle ces relations les contingences de renforcement. Or, il faut comprendre l'action des contingences de renforcement exactement comme l'action de la sélection naturelle : l'organisme présente a priori une variation de comportements ; ces comportements entraînent des conséquences dans l'environnement, conséquences qui vont, en fonction de leur effet sur l'organisme, sélectionner les comportements qui les ont produites. Ainsi, les comportements ayant des conséquences favorables pour l'organisme (les renforçateurs) se maintiennent dans le répertoire (renforcement) de celui-ci alors que les autres disparaissent. Selon Skinner, ce principe est suffisant pour rendre compte de l'émergence de tous les comportements y compris les plus complexes [DONAHOE, BURGOS PALMER 93]. Valider une hypothèse aussi forte nécessite l'utilisation de simulations informatiques afin d'étudier l'évolution d'un système fonctionnant selon des principes sélectionnistes d'où une convergence d'intérêt entre psychologues et informaticiens étudiant l'adaptation [DONAHOE, BURGOS ET PALMER 93].

Nous nous sommes placés dans ce cadre sélectionniste afin d'analyser de manière originale la dynamique des interactions sociales¹. Nous proposons qu'une interaction sociale est la résultante d'une contingence de groupe c'est-à-dire d'une contingence de renforcement dans laquelle les conséquences du comportement d'un individu ne dépendent pas seulement du comportement de cet individu mais aussi du comportement d'autres sujets [SCMITT 84]. Comme toute contingence, celle-ci va sélectionner directement les comportements contingents au renforçateur (appelons-les comportements principaux). Mais en faisant cela, elle va permettre à d'autres comportements (appelons-les comportements collatéraux) d'être renforcés parce qu'ils ont un effet en retour sur le comportement principal. Notamment, les comportements collatéraux d'un sujet x seront renforcés par leurs conséquences sur le comportement principal d'un sujet y. Nous avons appliqué ce schéma d'analyse au dilemme du prisonnier [AXELROD 92]. Il s'agit d'une caractéristique fondamentale de la plupart des situations de coopération réelles. Il traduit le fait que les situations de coopération ne renforcent pas seulement les comportements coopératifs mais aussi l'exploitation des sujets coopérants. Opérationnellement, dans le cadre d'une interaction coopérative entre deux sujets, cela se traduit par le fait que les gains de la coopération mutuelle sont supérieurs à ceux de la non coopération (trahison) mutuelle mais, si l'un des sujets coopère tandis que l'autre ne coopère pas, alors ce dernier reçoit le gain maximum et l'autre reçoit le gain minimum². Cela peut menacer le maintien de la coopération comme cela a été démontré en laboratoire et dans des environnements plus naturels. Pourtant, certains organismes ont réussi à éviter le piège de dilemme,

¹ Cette hypothèse s'inspire d'une approche sélectionniste de l'évolution des cultures connue en anthropologie sous le nom de matérialisme culturel. Voir [HARRIS 79] pour plus de détails.

² Il existe une dernière condition pour qu'une matrice 2*2 puisse être considéré comme un dilemme du prisonnier mais nous n'en parlerons pas ici.

l'exemple le plus connu étant celui des chauve-souris vampires qui régurgitent du sang pour maintenir en vie leurs congénères rentrant bredouille de leurs chasses nocturnes [DAWKINS 90]. La question est donc de savoir comment certains systèmes arrivent à maintenir les comportements coopératifs. Les chercheurs ont traditionnellement adopté une approche "écologique". Ils ont considéré que les agents interagissaient au sein d'une population et utilisaient des stratégies. Le succès d'un agent dans le dilemme étant censé refléter son succès reproductif, les chercheurs ont étudié la manière dont la proportion d'une stratégie évoluait au sein d'une population, cherchant à voir si une stratégie prenait le dessus [AXELROD 92] ou si des phénomènes dynamiques plus complexes pouvaient se manifester [DELAHAYE ET MATHIEU 95]. Une telle approche laisse dans l'ombre l'origine des stratégies et la manière dont elles s'implémentent dans le comportement des organismes. De plus, elle postule que la solution au dilemme ne peut être que d'origine phylogénétique. Enfin, elle suppose que le dilemme a donné lieu à l'évolution de comportements ou de structures particulières et donc que les capacités d'apprentissage des organismes vivants étaient incapables de résoudre ce problème. Notre recherche se base sur l'hypothèse contraire et suppose donc que les mécanismes d'apprentissage par renforcement sont suffisants pour permettre à des organismes n'utilisant pas de règles de résoudre le dilemme. Des études de la coopération en général, insistant aussi sur le rôle du renforcement, ont déjà été menées par certains chercheurs [DARCHEVILLE 91] [SCHMITT 84]. Selon notre hypothèse, le principe de sélection par les conséquences est suffisant pour comprendre comment un organisme peut apprendre à coopérer dans le cadre des interactions avec ses congénères. Cela n'est possible que si on considère que les organismes ne disposent pas seulement dans leur répertoire de comportements de coopération ou de non-coopération mais aussi d'autres comportements. Certaines recherches précédentes sur le dilemme aurait donc posé le problème de manière trop restrictive en limitant de manière excessive les comportements pouvant être émis par les agents. [ITO 96] avait également étudié la manière dont des agents autonomes pouvaient découvrir des solutions au dilemme. Néanmoins, ses agents utilisaient des stratégies définies par des algorithmes. Ces algorithmes pouvaient être modifiés de manière plus ou moins aléatoire selon un ensemble de règles précises. De plus, les conditions dans lesquelles ces règles se modifiaient nécessitaient qu'un agent dispose d'une quantité d'informations considérable sur le comportement des autres agents. La philosophie sous-jacente est donc très différente de celle de la présente recherche. Par ailleurs, les agents de Ito n'étaient pas basés sur des recherches empiriques. Enfin, Ito étudiait l'évolution des stratégies au sein de populations, chose qui n'a pas été faite ici.

Dans la suite, nous considérerons le dilemme comme une contingence de groupe sélectionnant les comportements de coopération et de non coopération mais permettant aussi le renforcement indirect de comportements collatéraux. En particulier, tout comportement d'un sujet A qui augmente la probabilité que le sujet B coopère sera renforcé. En effet, dans un dilemme du prisonnier, quoi que l'on fasse, il est toujours préférable que l'autre coopère. C'est l'émergence de ces comportements de contrôle qui va expliquer que certains groupes ne tomberont pas dans le piège de la non coopération mais convergeront, en fonction de la dynamique de l'interaction, soit vers la coopération (les deux sujets coopèrent et se contrôlent mutuellement), soit vers l'exploitation (un seul sujet coopère. L'autre ne coopère jamais mais émet constamment des comportements de contrôle). Nous avons testé cette hypothèse dans une expérience réalisée en collaboration avec Samuel Delepoulle avec des sujets humains [JOZEFOWIEZ 97] et dans une simulation informatique utilisant des agents virtuels que nous allons exposer ici.

Description des agents.

Les agents sélectionnistes que nous avons conçus sont construits en mettant en relation deux équations proposées dans la littérature psychologique pour formuler les processus de base se manifestant lors des

apprentissages. Ces équations sont donc soutenues par des faits empiriques et ont été développées pour rendre compte de phénomènes n'ayant rien à voir avec la coopération.

Le première équation qui gouverne notre système est l'équation de Staddon-Zhang [STADDON ET ZHANG 90]. Cette équation a été formulée afin d'expliquer comment un renforçateur pouvait augmenter la fréquence d'émission du comportement qui lui est contingent. On considère que l'organisme dispose de différents répertoires comportementaux, un répertoire comportemental étant défini comme un ensemble de comportements mutuellement incompatibles. Chaque comportement (ou unité comportementale) au sein d'un répertoire possède un niveau d'activation qui est recalculé à chaque instant t par l'équation (1) :

$$a_i(t+1) = \theta(t).a_i(t) + \varepsilon.(1 - \mu) \quad (1)$$

où

$a_i(t)$ est le niveau d'activation du comportement i à l'instant t ,

ε est une variable aléatoire distribuée uniformément sur $[0,1]$. ε est calculé indépendamment pour chaque unité à chaque instant t .

Le comportement ayant le niveau d'activation le plus élevé dans le répertoire est émis. Si l'émission d'un comportement n'est pas suivie d'un renforçateur, alors

$$\theta(t) = \mu \quad (2)$$

où μ est un facteur de mémoire à court terme toujours inférieur à 1. Dans la simulation, $\mu = 0,4$. Par contre, si le comportement est suivi d'un renforçateur, alors

$$\theta(t) = \mu + \gamma(t) \quad (3)$$

où $\gamma(t)$ est un signal d'activation général calculé par l'équation (4) et qui dépend de l'intensité du renforcement et de l'histoire de l'agent. Ce signal est délivré à toutes les unités comportementales d'un répertoire, pas seulement à celle correspondant au comportement qui vient d'être émis et qui est responsable de l'apparition du renforçateur. Comme l'équation (3) magnifie les différences entre les niveaux d'activation des unités comportementales, elle va favoriser le comportement renforcé dans sa compétition avec les autres comportements du répertoire. Le comportement renforcé va donc être émis plus souvent que les autres comportements même s'il ne sera pas émis systématiquement, le comportement de l'agent gardant une certaine variabilité.

La seconde équation que nous avons utilisée est celle de Rescorla-Wagner [RESCORLA ET WAGNER 72]. Cette équation est fréquemment utilisée dans la recherche psychologique sur l'apprentissage et permet de rendre compte de nombreux faits expérimentaux. Elle est identique à l'équation de Widroff-Hoff et est à la base du développement de certaines méthodes d'apprentissage par renforcement dans le domaine des réseaux de neurones artificiels [SUTTON ET BARTO 90]. Elle repose sur l'idée que l'organisme "cherche" à prédire ce qui va se passer par la suite, à savoir l'occurrence d'un renforçateur et l'intensité de celui-ci. Il n'y a apprentissage que lorsque ces attentes sont violées. Ce processus est résumé dans les équations (4) et (5) qui implémentent l'équation de Rescorla Wagner dans le cadre de nos agents :

$$\gamma(t) = E(t) + \rho.[\lambda(t) - E(t)] \quad (4)$$

$$E(t) = \sum_{i=1}^{n_j} w_i(t).x_i(t) \quad (5)$$

où

$E(t)$ représente les attentes de l'agent concernant le renforcement à l'instant t ,

$\lambda(t)$ est la quantité de renforçateur effectivement reçue par l'agent à l'instant t ,

ρ est une constante réglant la vitesse avec laquelle l'agent modifie ses attentes ($\rho=0.1$ dans la simulation),

$x_i(t)$ est un paramètre associé au comportement i du répertoire comportemental. $x_i(t)=1$ si le comportement qui lui est associé vient d'être émis, 0 sinon.

$w_i(t)$ est l'attente associée au comportement i concernant le renforcement. Au début de la simulation, $w_i(0) = 0$.

n_j est le nombre total d'unités comportementales du répertoire j .

$\gamma(t)$ active ensuite le répertoire comportemental comme décrit dans l'équation (3). Mais, dans le cas où les attentes de l'agent sont violées, $\gamma(t)$ permet également de modifier ses attentes comme décrit dans les équation (6) et (7) où on a

$$w_i(t+1) = \gamma(t) \quad (6)$$

si $x_i(t) \neq 0$ et

$$w_i(t+1) = w_i(t) \quad (7)$$

sinon. Donc, si le comportement i vient d'être émis ($x_i(t) \neq 0$) et si les attentes de l'agent sont violées ($E(t) \neq \lambda(t)$) alors il y a modification des attentes ($w_i(t+1) \neq w_i(t)$). Dans tous les autres cas, les $w_i(t)$ restent inchangés. Dans le cas où l'agent dispose de plusieurs répertoires comportementaux, chaque répertoire dispose de son propre système de renforcement exploitant les équations (4), (5), (6) et (7). Les répertoires agissent quasiment de manière autonome, la seule chose qu'ils aient en commun étant la quantité de renforcement délivrée à chaque instant, $\lambda(t)$. Les agents sont implémentés via un programme simple écrit en Turbo Pascal.

Simulation du dilemme du prisonnier.

Les agents que nous avons utilisés dans la simulation sont composés de deux répertoires comportementaux, un pour les comportements en rapport avec la coopération, l'autre pour les comportements en rapport avec le contrôle. Chacun comporte deux unités comportementales : l'une correspondant aux comportements de coopération et l'autre correspondant aux comportements de non coopération dans le premier répertoire ; l'une correspondant aux comportements de contrôle, l'autre pour des comportements n'ayant jamais aucune conséquence particulière dans le deuxième répertoire. Les comportements de contrôle sont définis comme étant des comportements dont l'émission par un sujet augmente les gains de l'autre sujet lorsque celui-ci coopère. Cette forme de contrôle était celle utilisée dans l'expérience de Jozefowicz et Delepoulle. D'autres formes de contrôle pourraient être envisagées. Il est probable que de tels comportements existent naturellement dans le répertoire de sujets réels, soit de manière innée (comme par exemple les expressions faciales ou les postures indiquant le contentement ou la menace chez de nombreux animaux), soit suite à un apprentissage (dans le cas de certains comportements sociaux par exemple). L'émergence de comportements collatéraux dans les interactions sociales est un fait courant [SCHMITT 84] Notre hypothèse est que ces comportements collatéraux sont des tentatives de contrôle d'autrui induites par la contingence de groupe.

La contingence de groupe dans laquelle les deux agents sont placés est décrite dans le tableau I qui donne les valeurs de $\lambda(t)$ pour un agent en fonction du comportement des deux agents. Le tableau I.A correspond à un dilemme du prisonnier classique. Le tableau I.B permet de voir l'effet des comportements de contrôle : l'émission de ceux-ci par un agent augmente légèrement les gains de la coopération pour l'autre agent.

Choix Agent 2 Choix Agent 1	Coopérer (A)	Ne pas coopérer (B)
Coopérer (A)	6 u	u
Ne pas coopérer (B)	10 u	2 u

tableau I.A : Gain de l'agent 1

Choix Agent2 Choix Agent1	Coopérer (A)	Ne pas coopérer (B)
Coopérer (A)	8 u	3 u
Ne pas coopérer (B)	10	2 u

tableau I.B : Gain de l'agent 1

Tableau I : Le tableau décrit la contingence de groupe utilisée dans la simulation. Le tableau I.A correspond à un dilemme du prisonnier classique. Le tableau I.B indique comment l'émission des comportements de contrôle par l'agent 2 modifie les gains de l'agent 1 dans la condition expérimentale. Dans la simulation que nous reportons ici, $u=0.08$.

Nous avons combiné les variables indépendantes suivantes selon un plan factoriel 2×2 :

- Le nombre d'itérations du système : chaque agent a été itéré 1000 ou 2000 fois. Le nombre d'itérations choisi correspond au temps nécessaire pour que le système se stabilise comme nous l'ont révélé des prétests menés uniquement avec l'équation de Staddon-Zhang.
- La possibilité de contrôler l'autre agent : dans la condition contrôle, les gains de chaque agent étaient calculés grâce au tableau I.A. Les comportements de contrôle n'avaient donc aucun effet. Dans la condition expérimentale, lorsqu'un agent émettait le comportement de contrôle, les gains de l'autre agent étaient calculés grâce au tableau I.B.

Dans chaque condition, on lance la simulation 20 fois. A chaque fois, on mesure dans chacun des couples d'agents le taux de coopération de chaque agent (c'est-à-dire le pourcentage d'essais durant lesquels il a émis les comportements de coopération) et leur taux de contrôle (c'est-à-dire le pourcentage d'essais durant lesquels l'agent a émis les comportements de contrôle). Puis, dans chaque condition, on comptabilise le nombre de couples ayant convergé vers la coopération, la non coopération et l'exploitation. Il y a deux critères de classification qui permettent de savoir comment catégoriser un couple : un critère strict et un critère lâche.

- Dans le cadre du critère lâche, on ne regarde que les taux de coopération. Si le taux de coopération des deux agents est supérieur à 50 %, le couple est considéré comme ayant convergé vers la

coopération. Si le taux de coopération des deux agents est inférieur à 50 %, le couple est considéré comme ayant convergé vers la non coopération. Si le taux de coopération d'un agent est supérieur à 50 % alors que celui de son partenaire est inférieur à 50 %, le couple est considéré comme ayant convergé vers l'exploitation.

- Dans le cadre du critère strict, on prend en compte à la fois les taux de coopération et les taux de contrôle : Si à la fois le taux de contrôle et le taux de coopération des deux agents sont supérieurs à 50 %, le couple est considéré comme ayant convergé vers la coopération. Si le taux de coopération de l'un des agents est supérieur à 50 % alors que celui de son partenaire est inférieur à 50 % et si de plus, le taux de contrôle de ce dernier agent est supérieur à 50 %, alors le couple est considéré comme ayant convergé vers l'exploitation. Quant au critère de classification pour la non coopération, il n'a pas changé.

Les résultats sont résumés dans les figures 1 et 2.

- On note que dans le cadre de la situation contrôle, on retrouve le résultat classique du dilemme du prisonnier. Dans les figures 1a et 1c, la majorité des couples d'agents convergent vers la non coopération et, surtout, aucun couple ne converge vers la coopération, même lorsqu'on utilise le critère lâche. Dans les figures 2a et 2c, la majeure partie du nuage de points pour les comportements de coopération se trouve nettement dans la partie inférieure gauche qui correspond à un taux de coopération inférieur à 50 % chez les deux agents. Quant aux taux de contrôle, le nuage de points les concernant se trouve au centre du graphique ce qui traduit une évolution au hasard.
- Les choses changent radicalement dès qu'on offre aux agents la possibilité de se contrôler l'un l'autre (figures 1b, 1d, 2b, 2d). Les résultats étant plus évidents pour 2000 itérations, c'est cette condition que nous commenterons (une tendance certaine s'observe néanmoins après 1000 itérations). Dans cette condition expérimentale, on assiste à une réduction notable du nombre de couples convergeant vers la non coopération : ils ne sont plus que 4 sur 20 alors qu'ils étaient 16 sur 20 dans la condition contrôle correspondante. De plus, on assiste à une émergence massive de la coopération : alors qu'aucun couple n'avait présenté ce comportement dans la condition contrôle correspondante, 12 couples selon le critère strict et 13 selon le critère lâche convergent vers cet état dans la condition expérimentale. De plus, la comparaison des figures 1b et 1d nous indique que, si l'attraction de la coopération augmente avec le nombre d'itérations, la stabilité de la non coopération diminue. Ces données sont corroborées par l'étude du nuage de points (figure 2) qui indique que des taux très élevés de coopération sont atteints, ce qui signifie que la coopération est stable une fois qu'elle apparaît. En ce qui concerne les comportements de contrôle, l'examen de la figure 3d montre clairement un déplacement du nuage de points vers la partie supérieure droite du graphe, ce qui indique des taux de contrôle nettement supérieurs à 50 %. De fait, on note que des taux particulièrement élevés sont atteints.

Néanmoins, cette élévation dans le taux de contrôle pourrait être un artefact dû au fait que la contingence du tableau I.B biaise le fonctionnement aléatoire d'un répertoire alors même qu'aucun des comportements de ce répertoire n'est renforcé. Ce phénomène est connu en psychologie sous le nom de comportements superstitieux [SKINNER 48]. Nous avons donc procédé à 2000 itérations en utilisant uniquement la contingence du tableau I.B afin de vérifier cette hypothèse. Les résultats sont fournis dans la figure 3. On note que, en effet, la contingence du tableau I.B provoque le renforcement superstitieux des comportements du répertoire 2 des agents. Néanmoins, comme on pouvait s'y attendre, ce renforcement touche l'une ou l'autre des unités du répertoire selon les essais. De ce fait, la forme du nuage de points est différente de celle observée dans la condition expérimentale : les points se dispersent sur l'ensemble de l'espace et on ne retrouve pas la tendance linéaire de la figure 3d.

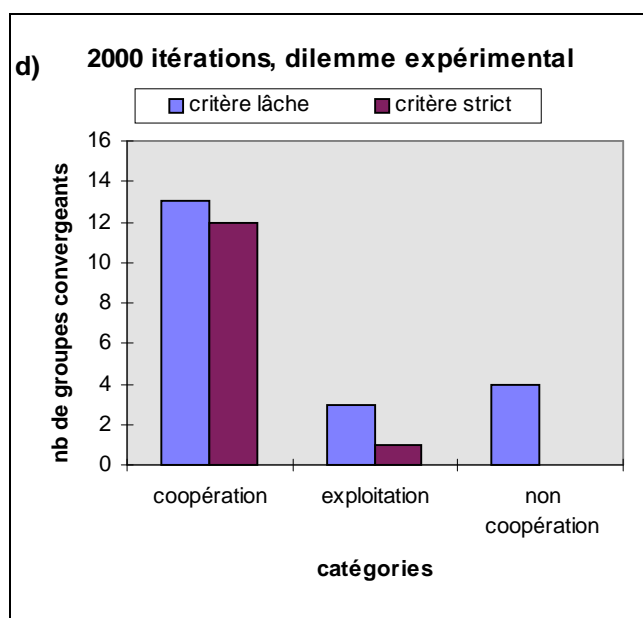
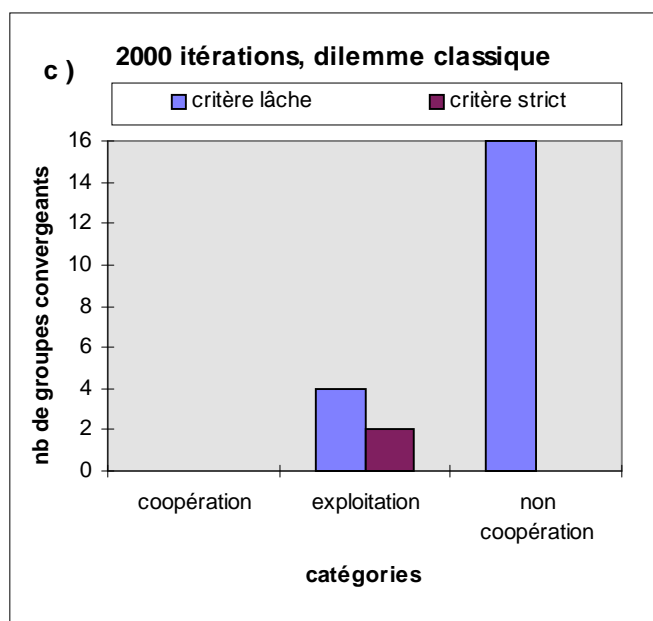
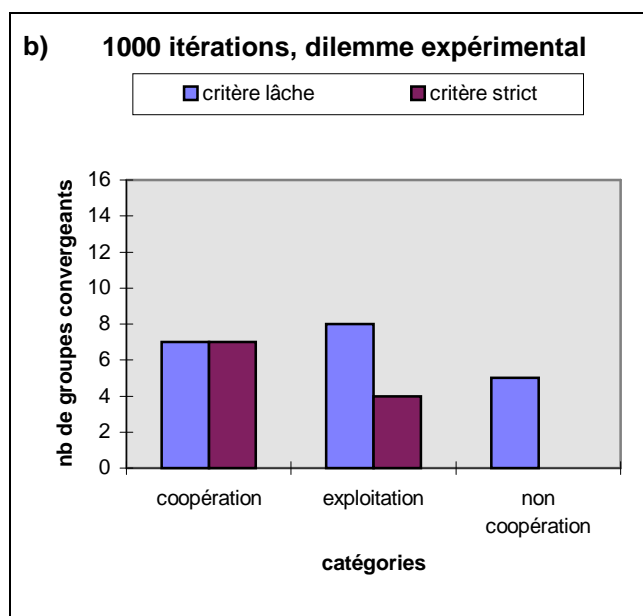
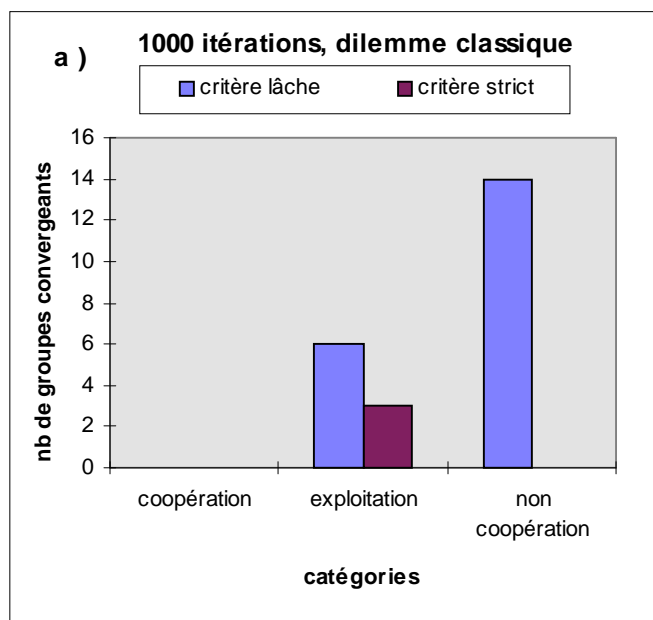
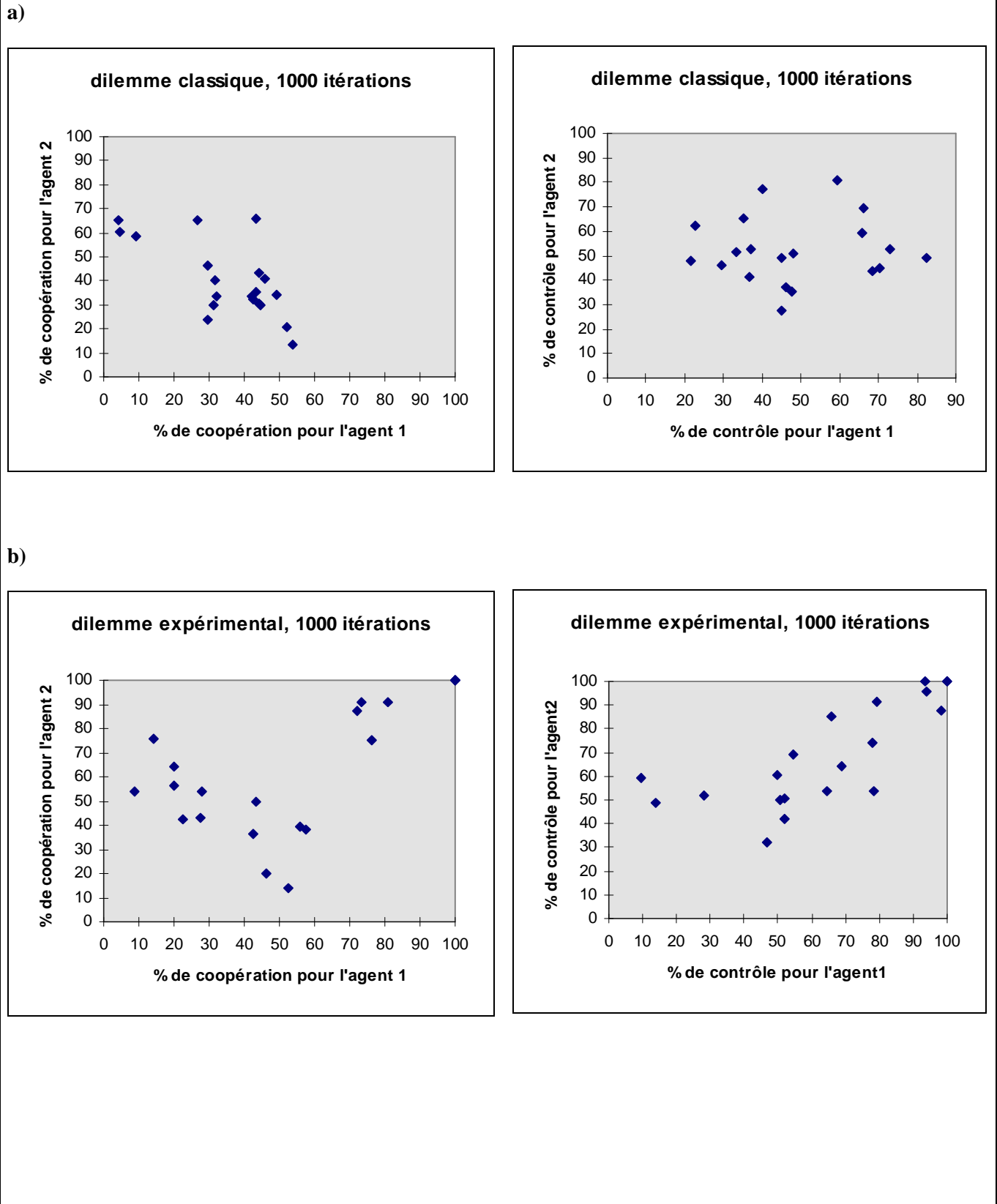


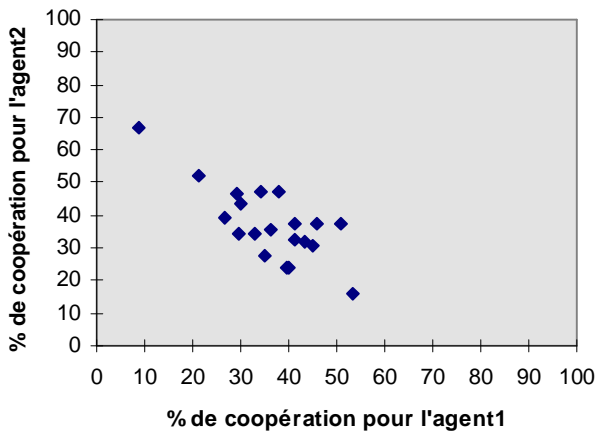
Figure 1 : convergence des couples d'agents en fonction de la condition .

Figure 2 : taux de coopération et de contrôle dans les couples d'agents en fonction de la condition (La figure se continue en page suivante).

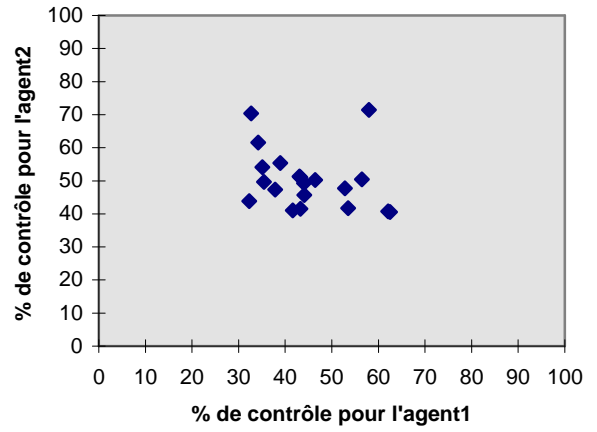


c)

dilemme classique, 2000 itérations

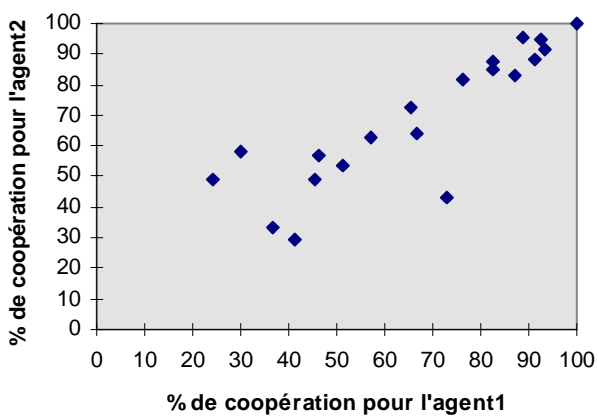


dilemme classique, 2000 itérations

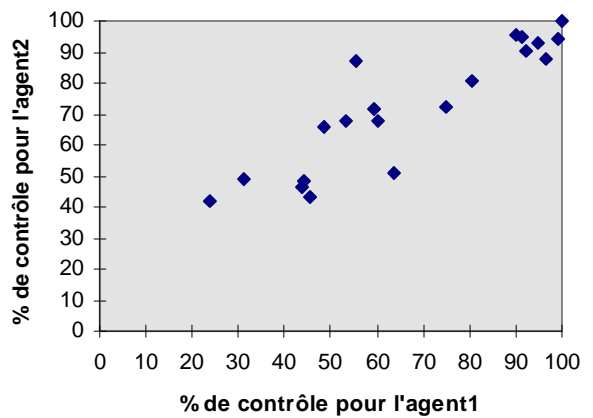


d)

dilemme expérimental, 2000 itérations



dilemme expérimental, 2000 itérations



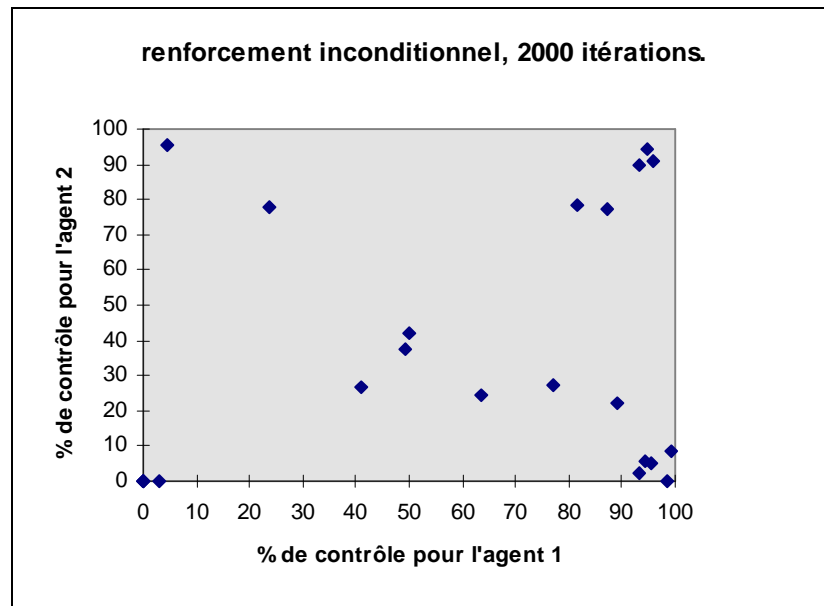


Figure 3 : taux de contrôle lorsque les agents sont soumis de manière inconditionnelle à la contingence du tableau I.B.

Conclusion.

En conclusion, on peut dire que les résultats vont dans le sens de nos hypothèses : la possibilité pour les agents de se contrôler l'un l'autre a modifié de manière radicale la dynamique de l'interaction et a permis aux agents d'éviter de tomber dans le piège du dilemme du prisonnier. Cela est dû à un renforcement indirect par la contingence de groupe des comportements de contrôle. Le succès de la simulation amène d'autres questions :

- On pourrait nous reprocher de n'avoir résolu qu'une variante du dilemme puisque la possibilité de contrôler l'autre change la donne. Nous pensons que cela n'est pas le cas, notre thèse étant justement que le dilemme du prisonnier ne se pose jamais sous sa forme "pure" aux sujets en raison de l'existence dans leur répertoire de comportements de contrôle.
- La réussite de la simulation est due à l'introduction de l'équation de Rescorla-Wagner. L'introduction de cette équation trouve sa justification dans une approche comportementale du problème dans laquelle la construction des agents est guidée par les connaissances actuelles sur la manière dont les organismes réels s'adaptent à leur environnement. De ce fait, une explication formelle du succès de la simulation n'a pas encore été fournie, même si intuitivement on comprend que le rôle de l'équation de Rescorla-Wagner est d'introduire une différence entre l'amplitude réelle du renforçateur et son effet sur la dynamique du répertoire comportemental. L'autre propriété importante de la dynamique des agents est certainement le fait que l'équation (1) diverge vers l'infini lorsque $\theta(t) > 1$ et converge vers 0 lorsque $\theta(t) < 1$. Dans le premier cas, les différences entre les niveaux d'activation des unités sont magnifiées de manière beaucoup plus importante que dans le deuxième cas et le comportement de l'agent devient dès lors plus inflexible. Nous pensons pouvoir arriver à une compréhension plus fine

des résultats de la simulation en réécrivant l'équation (6) de façon à pouvoir exprimer $w_i(t)$ à chaque instant t en fonction de l'amplitude et de la fréquence du renforçateur ainsi qu'en fonction de ρ . Cela permettrait de faire de nouvelles prédictions quant aux situations dans lesquelles la coopération ou la non coopération émergent. Ces prédictions pourraient donner lieu à des expériences avec des organismes réels.

- L'une des limites de ce travail est qu'il n'envisage qu'une forme de contrôle, forme qui de plus renforce spécifiquement la coopération. Ces comportements sont censés avoir été installés par la sélection naturelle mais il faudrait le démontrer. Il serait donc intéressant d'observer l'évolution des répertoires comportementaux d'une population d'agents, ces répertoires pouvant être constitués de plusieurs types de comportements de contrôle dont certains ne renforceraient pas la coopération.

Références.

[AXELROD 92] AXELROD, R. *Donnant, donnant : théorie des comportements coopératifs*,. Odile Jacob, 1992.

[DARCHEVILLE 91] DARCHEVILLE, J. C. *Une approche opérante de la règle : apprentissage de la règle d'intersection de deux ensembles et de la multiplication logique*. science et comportement, 21(2), 80-102, 1991.

[DAWKINS 90] DAWKINS, R. *Le gène égoïste* (seconde édition), Odile Jacob, 1990.

[DELAHAYE ET MATHIEU 95] DELAHAYE, J. P. & MATHIEU, P. *Complex strategies in the iterated prisoner's dilemma*, in A. ALBERT (Ed), *chaos and society*,. IOS Press, 1995.

[DONAHOE, BURGOS ET PALMER 93] DONAHOE, J. W., BURGOS, J. E., & PALMER, D. C. *Selectionist approach to reinforcement*, journal of the experimental analysis of behavior, 60, 17-40, 1993.

[EDELMAN 92] EDELMAN, G. M. *Biologie de la conscience*, Odile Jacob, 1992.

[HARRIS 79] HARRIS, M. *Cultural materialism*, Random House, 1979.

[ITO 96] ITO, A. *How do selfish agents learn to cooperate ?* proc. ALIFE V, MIT Press, 1996.

[JOZEFOWIEZ 97] JOZEFOWIEZ, J. *Proposition pour une analyse sélectionniste des interactions sociales : l'exemple de la coopération*, mémoire non-publié de maîtrise, Lille, 1997.

[RESCORLA ET WAGNER 72] RESCORLA, R. A. & WAGNER, A. R. *A theory of pavlovian conditioning : variations in the effectiveness of reinforcement and nonreinforcement*, in A. H. BLACK & W. F. PROKASY (Eds), *classical conditioning II*, Prentice-Hall, 1972.

[SCHMITT 84] SCHMITT, D. R. *Interpersonal relations : cooperation and competition*,. journal of the experimental analysis of behavior, 42, 377-383, 1984.

[SKINNER 81] SKINNER, B. F. *Selection by consequences*, science, 213, 501-504, 1981.

[STADDON ET ZHANG 93] STADDON, J. E. R & ZHANG, Y. *On the assignment-of-credit problem in operant learning*, in M. L. COMMONS, S. GROSSBERG & J. E. R. STADDON (Eds), *Neuronal networks of conditioning and action*, Erlbaum, 1993.

[SUTTON ET BARTO 90] SUTTON, R. S. & BARTO, A. G. *Time derivative models of pavlovian reinforcement*, in M. GABRIEL & J. MOORE (Eds), *learning and computational neuroscience : foundations of adaptive networks*, MIT Press, 1981.

