



HAL
open science

Méta-Stratégies pour le Dilemme Itéré du Prisonnier

Jean-Paul Delahaye, Philippe Mathieu

► **To cite this version:**

Jean-Paul Delahaye, Philippe Mathieu. Méta-Stratégies pour le Dilemme Itéré du Prisonnier. 24e Journées Francophones sur les Systèmes Multi-Agents (JFSMA'16), Oct 2016, Rouen, France. pp.13-22. hal-01378567

HAL Id: hal-01378567

<https://inria.hal.science/hal-01378567>

Submitted on 10 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Méta-Stratégies pour le Dilemme Itéré du Prisonnier

J.P. Delahaye^a P. Mathieu^a
jean-paul.delahaye@univ-lille.fr philippe.mathieu@univ-lille.fr

^aCRIStAL Lab, UMR 9189 CNRS, Université des Sciences et Technologies de Lille, France

Résumé

Nous présentons deux méthodes basées sur des notions différentes de méta-stratégies, permettant d'évaluer et de différencier les comportements généraux de coopération et d'agressivité des stratégies au dilemme itéré du prisonnier (IPD). La première méthode consiste à prendre des classes de stratégies comportant parfois plusieurs milliers d'éléments (technique des "classes complètes"), à regrouper les stratégies en sous-classes selon leurs "tempéraments" coopératifs ou agressifs puis à évaluer globalement ces sous-classes. Quatre types de comportements sont confrontés et comparés. La seconde méthode consiste à opérer des transformations, sortes de méta-stratégies, qui "adoucissent" ou "endurcissent" les stratégies d'une classe fixée et à comparer la classe initiale avec les classes transformées. Bien que les classes de stratégies initiales envisagées soient variées, les résultats obtenus sont remarquablement stables et convergents. Les conclusions obtenues qui permettent des commentaires fins sur les divers types de coopération et d'agressivité ont donc une valeur générale. Les méthodes sont ainsi assimilables à des preuves expérimentales.

Mots-clés : *Theorie des jeux, Dilemme du prisonnier, stratégies d'agents, comportement*

Abstract

We present in this paper two methods to evaluate and differentiate the general behaviours of cooperation and aggressiveness of strategies in the iterated prisoner's dilemma (IPD). The first method involves taking classes of strategies, sometimes with thousands of items ("complete classes" technique), grouping strategies into subclasses based on their cooperative or aggressive "temperament" then comprehensively assess these subclasses. Four kind of behaviour are confronted and compared. The second method is to operate transformations that "soften" or "harden" strategies of a given set and compare them with the results obtained with the initial set. Although examined classes of initial strategies are different, the results are remarkably stable and convergent. The conclusions that

allow us to make precise comments about several kind of kindness and aggressiveness, have therefore a general scope. These methods can be then assimilated to experimental proofs.

Keywords: *Game Theory, Iterated prisoner's Dilemma, Agent's Strategy, Behaviour*

1 Introduction

Le dilemme itéré du prisonnier est un modèle mathématique et informatique permettant d'évaluer les choix stratégiques envisageables dans le monde biologique, social, politique ou économique [3, 14, 18, 20, 21, 24, 26]. Depuis quelques années, un renouveau d'intérêt s'est manifesté pour cet outil de compréhension des interactions coopératives [1, 2, 7, 9, 10, 11, 13, 23, 27, 28].

Une des conclusions générales tirées des simulations de grands ensembles d'individus jouant à ce jeu est qu'il vaut mieux être coopératif qu'agressif. En particulier l'attitude en début de jeu est considérée comme déterminante [3, 25, 29]. Cependant, cette conclusion est un peu vague et d'ailleurs elle n'est pas vraie d'une façon absolue : la stratégie la plus coopérative `all_c` ne réussit pas très bien. Une analyse plus fine et rigoureuse est nécessaire pour savoir quand, et de quelle façon la coopération est souhaitable.

Nous avons conçu deux méthodes basées sur des notions de méta-stratégies pour étudier cette question et produire des résultats stables et convergents assimilables à des preuves expérimentales.

Pour la première méthode nous séparons diverses classes de stratégies en sous-classes selon leurs "tempéraments" et nous étudions les résultats moyens des classes. La variété et l'étendue des classes (comportant jusqu'à 5000 stratégies) que nous avons soumises à cette procédure, et les résultats convergents obtenus donnent des indications précises et fiables sur les comportements souhaitables ou à éviter.

Pour la seconde méthode, nous utilisons des *transformations d'ensembles massifs de stratégies*. Nous définissons des procédés généraux modifiant légèrement tous les éléments de ces ensembles de stratégies qui les rendent plus coopératives ou plus agressives ("adoucissement" et "endurcissement"). Une large variété de procédés concurrents est mise en compétition. Nous mesurons l'effet de ces transformations à l'aide de confrontations généralisées entre stratégies dans des arènes évolutives. À nouveau, les résultats obtenus sont stables et indiquent donc quelles sont les transformations bénéfiques ou pénalisantes, et donc quels sont les traits de caractère souhaitables ou à éviter. La robustesse des résultats permet d'affiner et de compléter les conclusions de la première étape. L'introduction d'un paramètre dans cette deuxième méthode crée une hiérarchie infinie d'adoucissements et d'endurcissements et conduit à la découverte de nouvelles stratégies performantes et donc particulièrement intéressantes car fondées sur des idées simples et clairement identifiables.

2 Le jeu, les tournois, les compétitions écologiques

Deux entités sont soumises au dilemme du prisonnier si chacune doit choisir entre coopérer (c) et trahir (d pour "Defect"). Elles sont rétribuées par R points si chacune joue c, par P points si chacune joue d, et reçoivent respectivement T et S points si l'une joue d et l'autre c. On écrit cela : $[c, c] \rightarrow R + R$, $[d, d] \rightarrow P + P$, $[d, c] \rightarrow T + S$. Pour que la situation soit celle d'un dilemme, on impose que $T > R > P > S$ et $T + S < 2R$. Les valeurs classiques qui servent de référence sont $T = 5$, $R = 3$, $P = 1$, $S = 0$. Ce dilemme devient intéressant lorsqu'il est itéré [3].

Exemple de face à face. La stratégie *tit_for_tat* commence par coopérer, puis joue au coup n ce que l'adversaire a joué au coup $n - 1$. La stratégie *per_ccd* joue périodiquement c, c, d, c, c, d, ... Leur face-à-face dans une rencontre en 10 étapes donne :

| | | | | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|---|---|----|
| | 3 | 3 | 5 | 0 | 3 | 5 | 0 | 3 | 5 | 0 | 27 |
| per-ccd | c | c | d | c | c | d | c | c | d | c | |
| tft | c | c | c | d | c | c | d | c | c | d | |
| | 3 | 3 | 0 | 5 | 3 | 0 | 5 | 3 | 0 | 5 | 27 |

À la première étape *tit_for_tat* coopère comme *per_ccd*, $[c, c]$, chacune gagne

donc 3 points d'après les règles de rétribution $[c, c] \rightarrow 3 + 3$, $[d, d] \rightarrow 1 + 1$, $[d, c] \rightarrow 5 + 0$. La seconde étape est à nouveau $[c, c]$, la troisième $[c, d]$, etc. ce qui conduit à l'issue des 10 étapes de face-à-face à 27 points pour *tit_for_tat* et autant pour *per_ccd*.

Exemple de tournoi. La méthode la plus simple pour comparer des stratégies consiste à organiser des "tournois". Chaque stratégie qui y participe rencontre chaque autre (par exemple dans un face-à-face comportant 10 étapes), et cumule les points gagnés à chaque étape pour chaque rencontre. On classe alors les stratégies ayant participé en fonction du total des points récoltés.

| | all_c | all_d | tit_for_tat | per_ccd | Total |
|-------------|-------|-------|-------------|---------|-------|
| all_c | 30 | 0 | 30 | 21 | 81 |
| all_d | 50 | 10 | 14 | 38 | 112 |
| tit_for_tat | 30 | 9 | 30 | 27 | 96 |
| per_ccd | 36 | 3 | 27 | 24 | 90 |

Le classement donne ici que la stratégie *all_d* (qui trahit à chaque coup) est gagnante. La moins bonne est *all_c* (qui coopère toujours).

À côté de la classification des stratégies par tournois, il existe des méthodes de tests plus subtiles auxquelles ne réussissent que les stratégies robustes. Les "*compétitions écologiques*" sont l'une de ces méthodes [3, 5, 4, 6, 8, 17, 18]. On met plusieurs exemplaires de chaque stratégie à tester dans une arène virtuelle et on y organise un tournoi. En fonction des points gagnés lors de ce tournoi, on fait évoluer les effectifs des stratégies ce qui définit une seconde génération : le nouvel effectif d'une stratégie est proportionnel à la fois à l'effectif initial et au nombre de points gagnés dans le tournoi par chaque exemplaire de la stratégie. La seconde génération par la même méthode donne une troisième génération, etc. On maintient l'effectif total constant. Les stratégies gagnantes (celles dont les effectifs deviennent les plus grands) sont efficaces dans des arènes variables.

Exemple de Compétition écologique. On obtient des courbes d'effectifs FIGURE 1, où on observe que *tit_for_tat* s'impose magistralement alors que *all_d* (gagnante du tournoi) après une période de croissance se fait éliminer. Sur le long terme, *tit_for_tat* est la meilleure des quatre et elle est la seule à survivre avec un petit nombre de *all_c*. La leçon est claire : pour gagner sur le long terme, être le vainqueur du tournoi ne suffit pas ; si votre

force provient de l'exploitation des faibles, une fois qu'ils se seront épuisés vous perdrez.

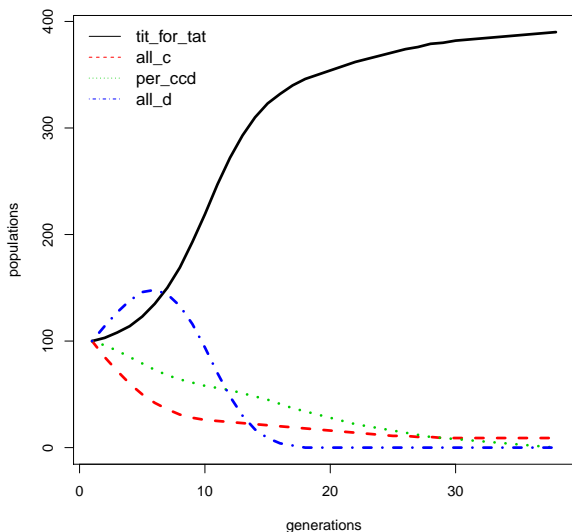


FIGURE 1 – Compétition écologique entre 4 stratégies

Les conclusions classiques sur le dilemme itéré sont les suivantes.

- Il n’y a pas de stratégies surclassant toute autre. Cependant certaines sont mauvaises dans pratiquement tous les environnements possibles (aussi bien en tournoi qu’en compétition écologique), alors que d’autres réussissent assez bien dans des environnements variés. C’est le cas de la célèbre stratégie *tit_for_tat*.
- Les stratégies efficaces possèdent souvent les caractéristiques suivantes [3, 14, 20, 21, 26].
 1. Elles sont réactives : elles répondent quand on les trahit, comme le fait *tit_for_tat*.
 2. Elles commencent par coopérer et face à un adversaire qui coopère, elles ne tentent pas de trahir.
 3. Elles savent être indulgentes : après une trahison de l’adversaire elles finissent par pardonner pour renouer la coopération.

Les compétitions écologiques confirment fréquemment les résultats obtenus en tournoi, mais en accroissent les contrastes. On y observe de plus un phénomène étonnant : sauf dans des cas très rares, l’arène finit par ne contenir que

des stratégies qui ne prennent jamais l’initiative de trahir (nous les nommons “initialement coopératives”, IC). Au bout de quelques générations, l’arène est occupée par des stratégies qui ne jouent entre elles que des coups $[c, c]$. L’arène se trouve donc dans un état de coopération généralisée. Le phénomène est frappant : bien qu’il n’y ait pas d’autorité de contrôle et que la tentation de la trahison soit présente pour tous à chaque coup joué, le mécanisme de l’évolution conduit à l’élimination de toutes les stratégies qui succombent à cette tentation.

3 Méthodes des sous-classes

Le comportement en début de face-à-face est très important et ses effets méritent une étude spécifique. Nous dirons qu’une stratégie est “initialement coopérative” (IC) si elle ne prend jamais l’initiative de trahir : elle coopère au premier coup d’un face-à-face et continue de le faire tant que l’autre ne l’a pas trahi. Tester si une stratégie est IC est simple : face à la stratégie qui coopère toujours, elle coopère toujours elle aussi (et donc en particulier au premier coup d’un face-à-face). La stratégie *tit_for_tat* est IC. On considère de manière identique les stratégies “initialement agressives” (IA) : elles trahissent sans cesse, tant que l’autre n’a pas coopéré. La stratégie *mistrust* (je joue comme *tit_for_tat* sauf que je trahis au premier coup) est IA.

Pour mesurer les avantages ou inconvénients qu’il y a à adopter des comportements IC ou IA, on considère des ensembles de stratégies variées (des méta-stratégies), on les confronte en tournois et en compétitions écologiques, et on mesure la réussite moyenne des diverses sous-classes, c’est à dire des diverses méta-stratégies.

Les quatre sous-classes que nous considérons sont la classe des stratégies initialement coopératives IC ; la classe complémentaire à IC qui est celle des stratégies “spontanément agressives”, SA (toute stratégie est donc soit IC soit SA) ; la classe des stratégies initialement agressives IA ; la classe complémentaire à IA qui est celle des stratégies “spontanément coopératives” SC (toute stratégie est donc soit IA soit SC). On note que toute stratégie IA est SA et que toute stratégie IC est SC.

Considérons l’ensemble des 17 stratégies de l’expérience 1 de [18]. Cet ensemble contient

| | | | | |
|-------------|----|----|----|----|
| all_c | | IC | | SC |
| all_d | IA | | SA | |
| tit_for_tat | | IC | | SC |
| spiteful | | IC | | SC |
| soft_majo | | IC | | SC |
| hard_majo | IA | | SA | |
| per_ddc | | | SA | SC |
| per_ccd | | | SA | SC |
| mistrust | IA | | SA | |
| per_cd | | | SA | SC |
| pavlov | | IC | | SC |
| tf2t | | IC | | SC |
| hard_tft | | IC | | SC |
| slow_tft | | IC | | SC |
| gradual | | IC | | SC |
| prober | IA | | SA | |
| mem2 | | IC | | SC |

TABLE 1 – Classement des 17 stratégies classiques décrites dans [18]

à la fois les stratégies les plus élémentaires et les stratégies reconnues performantes par la littérature sur le dilemme itéré du prisonnier [3, 5, 8, 12, 15, 18, 19, 22, 30].

La simulation se fait avec des rencontres en face-à-face de 1000 étapes et en utilisant les valeurs classiques des paramètres $T = 5$, $R = 3$, $P = 1$, $S = 0$. Lors de la rencontre en tournoi (TABLE2) les stratégies SA gagnent en moyenne, viennent ensuite les SC, puis les IA et enfin les IC : $SA > SC > IA > IC$.

| Tournoi | | | Compet Ecolo | | |
|---------|----|-------|--------------|----|--------|
| Cat | Nb | Moy | Cat | Nb | Moy |
| SA | 7 | 41099 | IC | 8 | 152.25 |
| SC | 13 | 40327 | SC | 13 | 130.38 |
| IA | 4 | 40258 | SA | 9 | 53.0 |
| IC | 10 | 39759 | IA | 4 | 0.0 |

TABLE 2 – Moyennes des IA, IC, SC, SA sur les 17 stratégies de base décrites dans [18]

Cet ordre semble signifier que coopérer initialement n'est pas judicieux. Cependant la méthode du tournoi autorise une stratégie agressive à exploiter les stratégies coopératives et non réactives comme `all_c`. Or dans une compétition écologique les stratégies coopératives ou non réactives (par exemple `all_c`) se font rapidement éliminer, ce qui entraîne alors l'élimination des stratégies trop brutales (comme `all_d`) qui exploitaient les premières. Le classement final obtenu lors de la compétition écologique met cela en évidence. Avec les 17 stratégies, il donne l'ordre suivant (TABLE2) entre les 4 sous-classes : $IC > SC > SA > IA$.

Pour réussir dans une compétition écologique,

il faut être efficace avec les stratégies efficaces, et non pas se contenter de profiter des "faibles". Il faut aussi *bien jouer contre soi-même*. Les compétitions écologiques sont plus réalistes que les tournois et cela produit un effet radical. Le meilleur type de comportement est la coopération initiale, et le pire l'agressivité initiale.

Pour s'assurer de la robustesse de cette conclusion, nous avons refait la même expérience avec d'autres classes de stratégies. En particulier nous avons considéré une classe de 1024 stratégies choisies de manière systématique sans aucun biais subjectif possible. Un tel biais, même inconscient, ne pouvait être exclu avec les 17 envisagées de la première expérience. Nous avons donc pris toutes les stratégies S dont le comportement dépend du dernier coup de S et des deux derniers coups joués par l'adversaire. Cette classe de stratégies se nomme $mem(1, 2)$; elle correspond à toutes les façons possibles dont peut jouer une stratégie déterministe mémorisant uniquement son dernier coup et les deux derniers coups de l'adversaire.

La FIGURE 2 montre pour quatre classes différentes comment évoluent les moyennes des effectifs des quatre sous-classes en fonction des numéros des générations. On voit qu'au départ (pour les premières générations) l'agressivité est une propriété efficace. Cependant plus le temps passe, laissant le mécanisme de la sélection naturelle s'opérer, plus les stratégies ayant de bonnes capacités à coopérer prennent le dessus.

D'une classe de stratégies à une autre, les dynamiques évolutives sont différentes (et sont par exemple plus ou moins longues) mais toutes arrivent au final à une stabilisation (la coopération généralisée) où l'ordre de domination est : $IC > SC > IA = SA$. Cet ordre doit donc être considéré comme universel dans le monde du dilemme itéré du prisonnier.

Dans la Figure 2 nous envisageons diverses classes de stratégies intéressantes. Successivement on considère les classes suivantes. (a) La classe des 17 stratégies sélectionnées dans [18]. (b) La classe des 32 stratégies déterministes qui jouent en fonction du dernier échange : stratégie à mémoire de 1 coup de soi-même et 1 coup de l'adversaire, notée $mem(1, 1)$. (c) La classe des 1024 stratégies déterministes qui jouent en fonction du dernier coup de soi-même et des deux derniers coups de l'adversaire, notée $mem(1, 2)$. (d) La classe des 2592 stratégies probabilistes à mémoire de 1 coup de

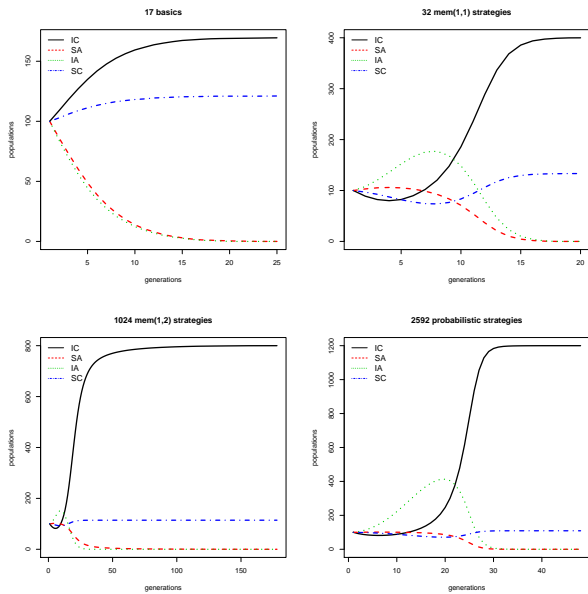


FIGURE 2 – Moyennes en fonction des numéros de générations des stratégies IC,SC,IA,SA pour quatre classes de stratégies

soi-même et de 1 coup de l'adversaire en envisageant de prendre pour probabilité possible tous les nombres $0, 1/5, 2/5, 3/5, 4/5, 1$. Cette classe est appelée Proba(5).

Pour chacune de ces classes nous avons réalisé une compétition écologique en calculant la moyenne des effectifs dans chaque sous-classe IC, SC, IA et SA génération après génération. On observe que de manière générale et indépendante de la classe, la sous-classe gagnante est IC, puis SC, puis (plus ou moins regroupées) IA et SA : $IC > SC > IA = SA$. Il est intéressant de noter que la dynamique des courbes varie très sensiblement d'une classe à l'autre. Dans presque tous les cas, lors des premières générations IA et SA l'emportent. Cela ne se produit pas pour la courbe des 17 car en réalité la sélection des 17 n'introduit sans doute pas assez de stratégies "faibles" servant "d'aliments" au IA et SA. La famille des 17 résultant d'une série de choix délibérés ne constitue peut-être pas une classe aussi neutre que les autres. La stabilité des résultats et en particulier de l'ordre $IC > SC > IA = SA$ dès que les effets d'une simulation de la sélection naturelle se sont produits est assez inattendue. Elle prouve que cet ordre est en quelque sorte universel. Une formulation de la morale de ces expériences pourrait s'énoncer : si l'évolution dure peu, il est intéressant d'être agressif, mais si on pense aux générations futures et lointaines, adopter des com-

portements coopératifs est le seul choix raisonnable.

La méthode directe d'analyse par les sous-classes est concluante, cependant elle ne permet pas d'affiner les conclusions sur l'intérêt des comportements coopératifs ou agressifs qui peuvent être de différentes natures, par exemple ne pas concerner seulement les débuts de parties. La méthode ne permet pas non plus de découvrir de nouvelles stratégies performantes. C'est pourquoi nous avons introduit la méthode des transformations qui ne s'occupe pas directement des stratégies mais de la réussite améliorée ou dégradée quand on produit un changement systématique des comportements de toute une classe de stratégies.

4 Méthode des transformations

L'idée de la méthode des transformations est de partir d'une classe de stratégies intéressantes C , et d'appliquer une méta-stratégie (une transformation) systématique de toutes les stratégies de la classe, ce qui donne une classe C' . Si chaque stratégie de la classe C' possède une attitude de coopération ou d'agressivité différente de celle de la stratégie avant transformation, le résultat moyen de la classe C' comparé à celui de la classe C indiquera si le changement opéré par la méta-stratégie est (1) une amélioration de son efficacité ou (2) une détérioration. Cette méthode conduira à classer des idées, éventuellement très générales et abstraites, en fonction des effets de leur application.

Dans un premier temps, nous avons introduit deux méthodes d'adoucissement et deux méthodes d'endurcissement. Quand on applique une méthode d'adoucissement à une stratégie S , on obtient une stratégie identique à S ou moins agressive. De même, appliquer une méthode d'endurcissement conduit à des stratégies inchangées ou plus agressives.

L'adoucissement `Soften1` consiste à partir d'une stratégie S et à définir la stratégie transformée S' suivante : au premier coup d'un face-à-face, S' coopère, et elle continue de coopérer tant que l'adversaire n'a pas trahi ; dès que l'adversaire a trahi, S' se met à jouer comme S aurait joué. La stratégie S' a donc un comportement identique à celui de S , sauf qu'elle ne l'applique que lorsque l'adversaire a pris l'initiative de trahir. Bien sûr, si on part d'une stratégie S initialement coopérative, S' est identique à S .

L'endurcissement `Harden1` est symétrique à la

méthode `Soften1` : S' trahit tant que l'autre n'a pas coopéré, et se met à jouer comme S dès que l'autre a coopéré. Cette transformation rend initialement agressives les stratégies S qui ne le sont pas.

L'adoucissement `Soften2` permet d'explorer une autre façon de rendre plus coopérative une stratégie quelconque. Quand on l'applique à une stratégie S , on obtient une stratégie S' qui se comporte comme S à la nuance près que, si au coup $n - 1$ l'adversaire a coopéré, S' choisit alors de coopérer. En clair, même quand S veut trahir, S' renonce et coopère à la condition que l'adversaire ait coopéré au coup $n - 1$. S' est une version indulgente de S , elle est prête à pardonner (sur la base du coup $n - 1$) à tout adversaire qui se montre coopératif et cela quel que soit le passé plus ancien des interactions. L'endurcissement `harden2` est défini symétriquement.

Dans une première expérience, on prend les 17 stratégies de base de [18]. On applique les deux méthodes d'adoucissement et les deux méthodes d'endurcissement à ces 17 stratégies. On obtient un total de 5 fois 17 stratégies, soit 85 stratégies. Certaines sont identiques entre elles mais par souci d'homogénéité et pour mener plus facilement les calculs sans risque d'erreur on ne cherche pas à supprimer les doublons. On observe TABLE3 les comportements globaux, c'est-à-dire les résultats moyens des 17 de base, des 17 ayant subi `Soften1`, des 17 ayant subi `Harden1`, etc.

| Tournoi | | Compet Eco | |
|---------|--------|------------|------|
| Soften1 | 218064 | Soften1 | 159 |
| Soften2 | 211604 | Soften2 | 141 |
| Normal | 198252 | Normal | 98.9 |
| Harden2 | 191036 | Harden2 | 98.2 |
| Harden1 | 151543 | Harden1 | 0 |

TABLE 3 – Résultats des adoucissements et endurcissements des 17 de base

Les enseignements à en tirer sont clairs : (a) s'adoucir est une bonne idée, qui améliore l'efficacité moyenne des stratégies ; (b) `Harden1` est de toutes les transformations envisagée la plus catastrophique.

En résumé la leçon est que pour les rencontres sans évolution (ce qui correspond aux tournois) il est utile mais assez peu sensible (voir TABLE 3) de modifier son comportement en l'adoucissant ou en l'endurcissant ; en revanche pour les compétitions écologiques, où une forme de sé-

lection naturelle agit, il est extrêmement bénéfique de pratiquer `Soften1` et `Soften2`. De plus, dans tous les cas, il ne faut pas endurcir votre comportement, et surtout pas en début de face-à-face.

Cette conclusion doit bien sûr être lue avec prudence. À nouveau pour conforter cette analyse, nous avons réalisé des simulations avec des ensembles de stratégies constitués de manière systématique et ne pouvant donc être sujet d'aucun biais subjectif.

Une expérience impliquant 5120 stratégies a été menée. On fait jouer les unes contre les autres les 1024 stratégies de la classe $mem(1, 2)$, et leurs versions modifiées par les 4 transformations précédemment énumérées (TABLE 4).

| Tournoi | | Compet Ecolo | |
|---------|----------|--------------|-----|
| Normal | 12063091 | Soften1 | 285 |
| Soften1 | 12061629 | Soften2 | 122 |
| Harden1 | 12060268 | Harden2 | 58 |
| Soften2 | 9705680 | Normal | 33 |
| Harden2 | 9704287 | Harden1 | 0 |

TABLE 4 – Résultats des adoucissements et endurcissements de $mem(1,2)$

Le classement du tournoi montre qu'il est sujet à une grande instabilité d'une classe à l'autre. En revanche, le classement des compétitions écologiques est lui, très stable. Le seul changement est la permutation dans le classement des compétitions écologiques entre `Harden2` et `Normal`. Cette permutation montre que `Harden2` est parfois bon : renforcer sa dureté une fois passé le début d'un face-à-face est une option à considérer.

D'autres expériences confirment que dans un monde où l'évolution joue son rôle en renforçant les effectifs des stratégies qui réussissent au dépens de celles qui obtiennent des scores médiocres, alors les comportements coopératifs sont les meilleurs. Il est particulièrement important de bien initier les débuts de face-à-face en coopérant aussi systématiquement que possible dans les premiers échanges. De plus, dans le déroulement d'un face-à-face savoir être indulgent est globalement payant. Un renforcement de la sévérité à condition qu'il ne s'applique pas aux débuts des face-à-face est parfois utile.

Pour la FIGURE 3 comme pour la FIGURE 2, nous avons cherché à savoir ce que donnent les familles transformées non pas seulement à l'issue d'un tournoi ou d'une compétition

écologique, mais génération après génération. Nous obtenons 4 figures comportant chacune 5 courbes : une courbe pour la classe des stratégies utilisée C sans y toucher, une courbe pour la famille C transformée par $Soften1$, une courbe pour la classe C transformée par $Soften2$, une courbe pour la classe C transformée par $Harden1$, une courbe pour la classe C transformée par $Harden2$. Les classes C utilisées sont successivement. (a) La classe des 17 stratégies sélectionnées. (b) La classe des 32 stratégies déterministes de $mem(1, 1)$. (c) La classe des 1024 stratégies déterministes de $mem(1, 2)$. (d) La classe des 1250 stratégies probabilistes à mémoire de 1 coup de soi-même et de 1 coup de l'adversaire en envisageant de prendre pour probabilité possible tous les nombres 0, 1/4, 1/2, 3/4, 1. Cette classe est appelée $Proba(4)$.

La conclusion est claire : $Soften1$ (adoucissement en début de face-à-face) est souhaitable d'une manière générale et plus intéressant que $Soften2$ qui reste cependant presque toujours intéressant. $Harden2$ (adoucissement tout le long du face-à-face) est souvent utile, alors que l' $Harden1$ (adoucissement en début de face-à-face) est toujours mauvais. Il faut noter que $Soften2$ est conditionnel : on est indulgent avec ceux qui sont coopératifs dont on oublie plus facilement le manque de coopération passé. Il n'est pas incompatible avec $Harden2$ qui consiste à changer son comportement pour le rendre plus agressif dès lors que l'adversaire refuse de coopérer. Finalement, on observe que parmi les quatre transformations envisagées, le plus souvent, trois d'entre elles sont intéressantes : l'adoucissement en début de face-à-face et au cours du face-à-face et l'endurcissement au cours du face-à-face. On pourrait traduire cela par : abordez toujours un éventuel collaborateur avec indulgence, soyez tolérant à propos de son comportement passé s'il se met à coopérer, renforcer vos réactions punitives s'il se montre agressif, mais surtout ne commencez jamais par trop de sévérité et d'exigences.

5 La durée des face-à-face

La question de la longueur des face-à-face ne doit pas être confondue avec la question de nombre des générations lors d'une compétition écologique. La longueur des face-à-face F indique si lorsque deux stratégies se rencontrent et interagissent, elles le font longuement ou brièvement. Le cas extrême est celui d'une durée de face-à-face de 1 qui correspond à un dilemme

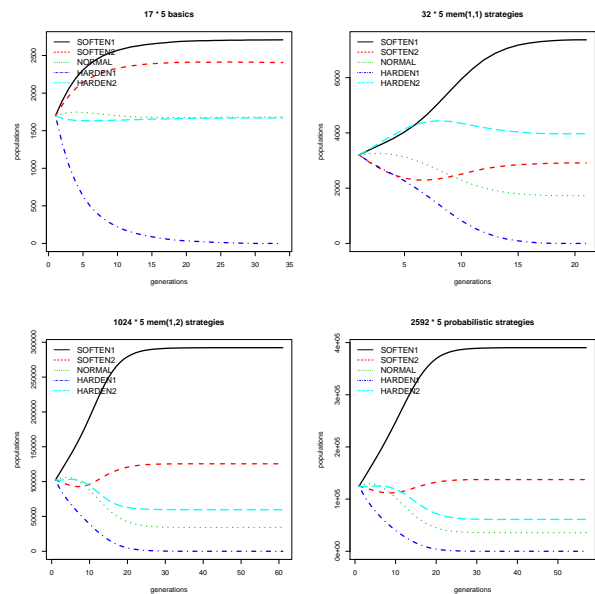


FIGURE 3 – Succès génération après génération des différentes classes transformées, et de la classe de départ, en considérant quatre classes de départ différentes.

du prisonnier non itéré. Cette durée F est un facteur important pour choisir sa stratégie et la modifier. Il est connu depuis le début des expériences de [3] que dès que la durée des face-à-face F dépasse 10 les résultats obtenus ne changent presque plus. En revanche pour des durées très courtes de face-à-face, les résultats peuvent sensiblement varier. C'est ce que la méthode des transformations prouve avec clarté, apportant des précisions sur les effets des courts face-à-face.

Pour étudier le rôle de F , nous avons dessiné 9 familles de courbes, correspondant à $F = 2, 3, \dots, 10$ en utilisant la méthode des transformations appliquée à la classe $mem(1, 1)$. Voir la FIGURE 4.

Ces résultats permettent de déterminer avec précision l'opportunité ou non selon les valeurs F des transformations d'adoucissement et d'endurcissement. Les observations sont les suivantes :

Quand $F = 2$, les endurcissements $Harden1$ et $Harden2$ sont les vainqueurs. Cela correspond simplement à l'idée que dans le cas d'une interaction courte entre stratégies, on est gagnant en trahissant, comme on l'est dans le cas d'un dilemme du prisonnier non itéré. Cette conclusion est valide même en cas d'évolution : être dans un monde soumis à la sélection naturelle

ne change pas l'intérêt d'un jeu agressif si tous les adversaires qu'on rencontre ne le sont que sur des durées très courtes.

Quand $F = 3$, la domination de Harden1 et de Harden2 devient moins nette, et l'avantage de Harden1 sur Harden2 ne se maintient pas.

Quand $F \geq 4$, Harden1 termine à la fin du classement. Soften1 qui est dominé dans les premières générations, termine largement en tête. Harden2 termine second après avoir dominé les premières générations. Soften2 et Normal (la classe sans transformation) après une histoire un peu compliquée terminent en position 3 et 4.

La stabilité globale des courbes et de leur position relative pour $F > 4$ établit comme on pouvait le prévoir que l'intérêt de la coopération qui est incertain lorsque F est petit devient net et stable dès que F dépasse 4 avec un subtil classement (dépendant des numéros de génération) qui prouve que Harden2 a lui aussi de l'intérêt. Être coopératif est une attitude recommandée, mais être réactif (comme le sont les familles résultant de Harden2 et de Soften2) est aussi un moyen de s'améliorer. Au final dès que F est assez grand et que l'évolution produit ses effets, au moins trois méthodes améliorent une classe donnée.

6 Paramètres d'adoucissement et d'endurcissement

Un autre type d'études a été mené pour évaluer l'efficacité des attitudes coopératives ou agressives dans le cadre du modèle du dilemme itéré du prisonnier avec la méthode des transformations. Cette méthode à paramètres permet de découvrir de nouvelles stratégies qui battent ou égalent les meilleures stratégies connues.

On introduit un paramètre k dans la définition d'une transformation. Il prend des valeurs entières 1, 2, 3, ...

Voici trois transformations paramétrées permettant d'adoucir une classe de stratégies, et trois transformations paramétrées permettant d'endurcir une classe.

- La transformation `Soften1-k` appliquée à la stratégie S, donne la stratégie S' dont le comportement est : tant que mon adversaire n'a pas trahi k fois au total je coopère, sinon je joue comme S l'aurait fait.

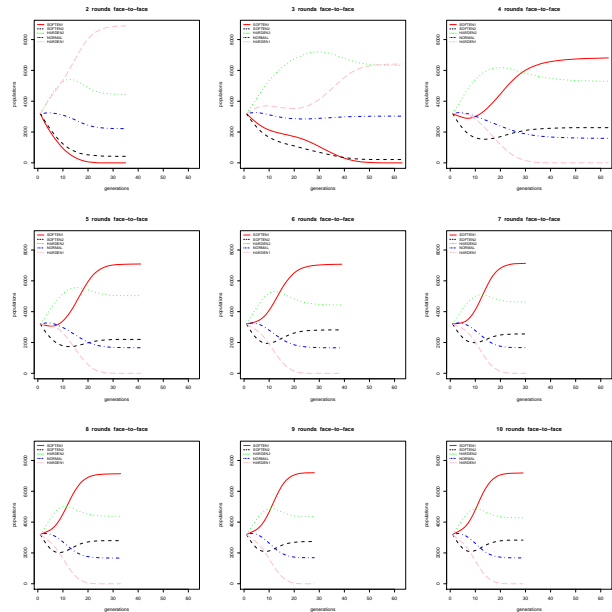


FIGURE 4 – Les 9 schémas de 5 courbes donnent les moyennes d'effectifs à partir de $mem(1, 1)$ pour des valeurs différentes des durées de face-à-face $F = 2, 3, \dots, 10$. Pour les petites valeurs de F , rendre son comportement plus agressif est payant même dans le cas d'un monde soumis à la sélection naturelle. Lorsque F dépasse 4, la domination des comportements agressifs cesse.

- La transformation `Harden1-k` est symétrique : tant que mon adversaire n'a pas coopéré k fois au total je trahis, sinon je joue comme S l'aurait fait.
- La transformation `Soften2-k` appliquée à la stratégie S, donne la stratégie S' dont le comportement est le suivant : si mon adversaire a coopéré au coup $n - 1$ OU au coup $n - 2$ OU ... OU au coup $n - k$ je coopère, sinon je joue comme S l'aurait fait.
- La transformation `Harden2-k` est symétrique : si mon adversaire a trahi au coup $n - 1$ OU au coup $n - 2$ OU ... OU au coup $n - k$ je trahis, sinon je joue comme S l'aurait fait.

En remplaçant les OU par des ET.

- La transformation `Soften3-k` appliquée à la stratégie S, donne la stratégie S' dont le comportement est le suivant : si mon adversaire a coopéré au coup $n - 1$ ET au coup $n - 2$ ET ... ET au coup $n - k$ je coopère, sinon je joue comme S l'aurait fait. Pendant les k pre-

- miers coups je suis certain de jouer S.
- La transformation Harden_{3-k} est symétrique : si mon adversaire a trahi au coup $n - 1$ ET au coup $n - 2$ ET ... ET au coup $n - k$ je trahis, sinon je joue comme S l'aurait fait. Pendant les k premiers coups je suis certain de jouer S.

La TABLE 5 fait découvrir que la stratégie `all_d` traitée par `Soften1` avec le paramètre 2 est meilleure que toutes les stratégies de l'expérience C de [18]. Cette expérience implique l'ensemble de 66 stratégies déterministes et probabilistes, que nous considérons comme la plus générale et la plus significative actuellement.

Le classement de la stratégie `all_d` transformée par `Soften1` pour les valeurs du paramètre k , passe de la place 50 (pour $k = 0$) à la place 6 (pour $k = 1$), puis devient la première pour $k = 2$. Ensuite elle est classée second (pour $k = 3$), quatrième (pour $k = 4$), et ainsi de suite.

Les 3 stratégies obtenues pour $k = 2, 3$ et 4 se révèlent performantes dans des contextes variés.

Les résultats montrent que devenir plus généreux et plus tolérant aux trahisons n'est payant que jusqu'à un certain point. Quand le paramètre d'adoucissement augmente, le bénéfice s'accroît un moment, puis passe par un maximum avant de diminuer. Il faut préférer les comportements coopératifs, oui mais pas trop quand même !

| K | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------------|----|----|----------|----|----|----|----|----|----|
| <code>all_d_Soft1</code> | 50 | 6 | 1 | 2 | 4 | 8 | 9 | 11 | 15 |
| <code>all_d_Soft2</code> | 51 | 32 | 24 | 25 | 27 | 29 | 28 | 31 | 33 |
| <code>all_d_Soft3</code> | 23 | 32 | 63 | 66 | 64 | 63 | 63 | 62 | 62 |

TABLE 5 – Classement des variantes de `all_d` dans l'ensemble des 63 stratégies de l'expérience C de [18]. On note qu'à $k = 2$ `all_d` traitée par `Soften1` devient la meilleure.

Les résultats obtenus sont tout à fait remarquables. La méthode d'adoucissement `Soften1`, en partant d'une stratégie agressive (la plus agressive possible !) donne (en choisissant la méthode d'adoucissement la meilleure et le paramètre d'adoucissement optimal) une stratégie meilleure que toutes celles identifiées comme bonnes dans la littérature.

Quand on prend ensuite les trois stratégies identifiées (car classées premières) et qu'on les place dans une soupe n'ayant aucun rapport avec celle utilisée pour les identifier (par exemple

`mem(1, 2)`) on obtient confirmation que les nouvelles stratégies identifiées sont excellentes comme on le peut le voir sur la FIGURE 5, respectivement avec les 17 de base, les `mem(1, 1)`, les `mem(1, 2)` et les `Proba(5)`.

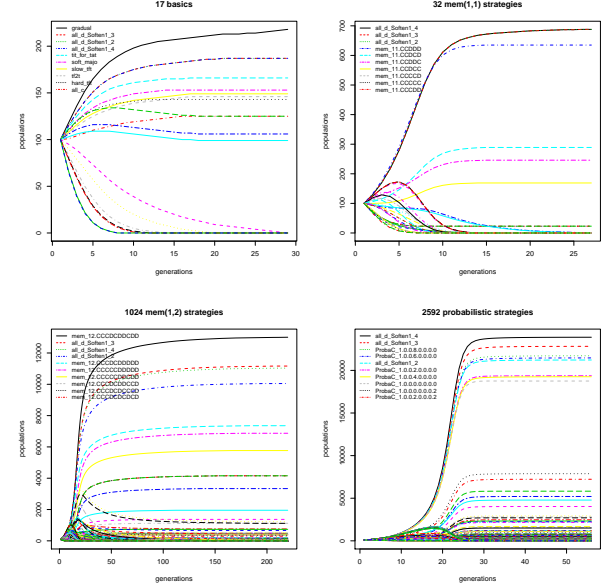


FIGURE 5 – Evaluation des 3 `all_d` adoucies dans des environnements différents

7 Conclusion

Dans ce travail nous montrons des résultats qui confirment et affinent des aspects généralement admis par la communauté, et d'autres qui contredisent ces jugements généraux. Nous avons proposé deux méthodes d'adoucissement qui ont des effets différents : la première, qui concerne les débuts de partie, se révèle sans surprise très bonne ; la seconde, qui concerne tout le déroulement de la partie, se révèle moins performante. Nous avons aussi proposé deux méthodes d'endurcissement symétriques aux méthodes d'adoucissement, et avec surprise, l'une des deux améliore globalement les comportements auxquels elle s'applique. Ceci contredit le "folk theorem" qui considère que l'agressivité est systématiquement pénalisante. Nous montrons pour finir que si on ajoute un paramètre d'adoucissement, il existe une valeur qui permet d'améliorer très significativement une stratégie y compris parmi les plus agressives (`all_d` par exemple). Il est remarquable qu'un modèle à la définition simple, après 30 ans d'études, produise encore des résultats nouveaux sur la façon dont il faut aborder les jeux de coopération dont l'importance est reconnue [16].

Références

- [1] Christoph Adami and Arend Hintze. Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. *Nature communications*, 4, 2013.
- [2] Christoph Adami and Arend Hintze. Corrigendum : Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. *Nature communications*, 5, 2014.
- [3] Robert Axelrod. The evolution of cooperation : revised edition. 1984, 2006.
- [4] Bruno Beaufils. *Modèles et simulations informatiques des problèmes de coopération entre agents*. PhD thesis, Université des Sciences et Technologie de Lille-Lille I, 2000.
- [5] Bruno Beaufils, Jean-Paul Delahaye, and Philippe Mathieu. Complete classes of strategies for the classical iterated prisoner's dilemma. In *Evolutionary Programming VII*, pages 33–41. Springer, 1998.
- [6] Bruno Beaufils and Philippe Mathieu. Cheating is not playing : Methodological issues of computational game theory. *Proceedings of ECAI 2006*, 141 :185, 2006.
- [7] Jean-Paul Delahaye. Le dilemme du prisonnier et l'illusion de l'extorsion. *Pour La Science, French version of Scientific American*, (435) :78–83, janv 2014.
- [8] Jean-Paul Delahaye, Philippe Mathieu, and Bruno Beaufils. The iterated lift dilemma. In *Computational conflicts*, pages 202–223. Springer, 2000.
- [9] Hao Dong, Rong Zhi-Hai, and Zhou Tao. Zero-determinant strategy : An underway revolution in game theory. *Chinese Physics B*, 23(7) :078905, 2014.
- [10] Christian Hilbe, Martin A Nowak, and Karl Sigmund. Evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences*, 110(17) :6913–6918, 2013.
- [11] Christian Hilbe, Torsten Röhl, and Manfred Milinski. Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nature communications*, 5, 2014.
- [12] Christian Hilbe, Arne Traulsen, and Karl Sigmund. Partners or rivals? strategies for the iterated prisoner's dilemma. *Games and Economic Behavior*, 2015.
- [13] Christian Hilbe, Bin Wu, Arne Traulsen, and Martin A Nowak. Evolutionary performance of zero-determinant strategies in multiplayer games. *Journal of theoretical biology*, 374 :115–124, 2015.
- [14] Graham Kendall, Xin Yao, and Siang Yew Chong. *The iterated prisoners' dilemma : 20 years on*. World Scientific Publishing Co., Inc., 2007.
- [15] Jiawei Li, Philip Hingston, and Graham Kendall. Engineering design of strategies for winning iterated prisoner's dilemma competitions. *Computational Intelligence and AI in Games, IEEE Transactions on*, 3(4) :348–360, 2011.
- [16] Curtis W Marean. The most invasive species of all. *Scientific American*, 313(2) :32–39, 2015.
- [17] Philippe Mathieu, Bruno Beaufils, and Jean-Paul Delahaye. Studies on dynamics in the classical iterated prisoner's dilemma with few strategies. In *Artificial Evolution*, pages 177–190. Springer, 2000.
- [18] Philippe Mathieu and Jean-Paul Delahaye. New winning strategies for the iterated prisoner's dilemma. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems AAMAS 15*, pages 1665–1666. International Foundation for Autonomous Agents and Multiagent Systems, 2015.
- [19] Shashi Mittal and Kalyanmoy Deb. Optimal strategies of the iterated prisoner's dilemma problem for multiple conflicting objectives. *Evolutionary Computation, IEEE Transactions on*, 13(3) :554–565, 2009.
- [20] MA Nowak. Evolutionary dynamics harvard university press. 2006.
- [21] Martin Nowak and Roger Highfield. *SuperCooperators : Altruism, evolution, and why we need each other to succeed*. Simon and Schuster, 2011.
- [22] Colm O'Riordan. A forgiving strategy for the iterated prisoner's dilemma. *Journal of Artificial Societies and Social Simulation*, 3(4) :56–58, 2000.
- [23] William H Press and Freeman J Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26) :10409–10413, 2012.
- [24] Anatol Rapoport and Albert M Chammah. *Prisoner's dilemma : A study in conflict and cooperation*, volume 165. University of Michigan press, 1965.
- [25] K Shigaki, Z Wang, J Tanimoto, and E Fukuda. Effect of initial fraction of cooperators on cooperative behavior in evolutionary prisoner's dilemma game. *PLoS one*, 8(11) :e76942–e76942, 2012.
- [26] Karl Sigmund. *The calculus of selfishness*. Princeton University Press, 2010.
- [27] Alexander J Stewart and Joshua B Plotkin. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences*, 110(38) :15348–15353, 2013.
- [28] Attila Szolnoki and Matjaž Perc. Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Scientific reports*, 4, 2014.
- [29] Jun Tanimoto. The impact of initial cooperation fraction on the evolutionary fate in a spatial prisoner's dilemma game. *Applied Mathematics and Computation*, 263 :171–188, 2015.
- [30] Claus Wedekind and Manfred Milinski. Human cooperation in the simultaneous and the alternating prisoner's dilemma : Pavlov versus generous tit-for-tat. *Proceedings of the National Academy of Sciences*, 93(7) :2686–2689, 1996.