LOGIQUE & CALCUL

Le dilemme du prisonnier et l'illusion de l'extorsion

Dans sa version itérée, le dilemme du prisonnier est un jeu qui aide à comprendre diverses situations sociales et stratégiques. Une controverse autour des stratégies d'extorsion a enrichi le débat.

Jean-Paul DELAHAYE

Le dilemme itéré du prisonnier est un jeu qui éclaire les comportements sociaux d'entités en concurrence sur un même terrain : organismes vivants se partageant une niche écologique, entreprises concurrentes se disputant un marché, personnes s'interrogeant sur l'intérêt de mener un travail en commun. Bien que le dilemme itéré du prisonnier soit fondé sur une simplification extrême, son étude mathématique reste difficile, et souvent seules les simulations réussissent à trancher les questions posées.

Cependant, un article paru en 2012 dans les Actes de l'Académie américaine des sciences, signé par le célèbre physicien de Princeton Freeman Dyson et par William Press de l'Université du Texas, a mis en évidence une famille de stratégies passées inaperçues. Les chercheurs démontraient mathématiquement certaines propriétés étonnantes de leurs « ZD-stratégies ». Celles-ci semblent capables d'imposer leur loi au jeu du dilemme itéré du prisonnier en pratiquant une forme d'extorsion. Cela étonna les spécialistes qui pensaient qu'aucune stratégie n'était « absolument » meilleure.

Le dilemme du prisonnier est celui auquel sont soumis deux agents ayant le choix entre coopérer (c) et trahir (t) et qui sont rétribués par R points chacun s'ils jouent tous deux c (coopère), par P points si chacun joue t (trahit), et qui reçoivent respectivement T et S points si l'un joue t et l'autre c. On peut résumer ces règles par : $[c, c] \to R+R$, $[t, t] \to P+P$, $[t, c] \to T+S$. On impose $T > R > P > S$ et on choisit souvent les valeurs : $T=5$, $R=3$, $P=1$, $S=0$, qui donnent : $[c, c] \to 3+3$, $[t, t] \to 1+1$, $[t, c] \to 5+0$.

Trahir ou coopérer ?

L'expression « dilemme du prisonnier » provient de l'histoire imaginaire de deux suspects arrêtés devant une banque auxquels un interrogateur essaie de faire avouer séparément qu'ils étaient sur le point de mener une attaque. Si l'un avoue, c'est-à-dire trahit son compère, et l'autre non, c'est-à-dire poursuit la coopération avec son compère (situation $[t, c]$), celui qui a avoué est libéré (rétribution de cinq années de liberté), celui qui n'a pas avoué va en prison cinq ans (maximum prévu pour une attaque de banque). Si les deux compères restent solidaires, $[c, c]$, ils vont deux ans en prison chacun pour port illégal d'arme (rétribution de trois années de liberté par rapport aux cinq ans du pire cas). Si les deux bandits se trahissent simultanément, c'est-à-dire avouent, $[t, t]$, ils font chacun quatre ans de prison (rétribution d'une année de liberté par rapport au pire cas en récompense de leurs aveux).

Dans cette situation, la trahison est logique : elle conduit toujours à un meilleur résultat que la coopération. En effet, si l'autre entité coopère, j'obtiens 5 points en

1. LES STRATÉGIES CLASSIQUES (A) et les gains obtenus (B et D) dans un tournoi où chacune rencontre pendant 1 000 coups chaque stratégie, y compris elle-même, calculés sur la base des points accordés à chaque coup (C). En totalisant les gains, on voit que les stratégies agressives sont mal classées. La stratégie *Méchante*, qui gagne ou fait jeu égal contre toute autre stratégie, est ainsi 10e sur 12. La stratégie *Donnant-donnant*, qui perd ou fait jeu égal contre toute autre, gagne pourtant le tournoi... car elle incite à la coopération.
Dans une simulation de l'évolution (E), les stratégies gagnant le plus de points sont celles qui ont le plus de descendants d'une génération à la suivante. On constate une convergence vers la coopération généralisée, et toutes les stratégies prenant l'initiative de trahir se font éliminer. Au bout de 30 générations, il ne reste que *Donnant-donnant, MajoMou, Rancunière, Donnant-donnant-dur, Gentille* dont les effectifs sont stables. Même les stratégies *Sondeur* et *Périodique-cct*, dont l'agressivité est pourtant modérée, sont éliminées.
Aux 12 stratégies précédentes, on ajoute les deux stratégies *Pavlov* (souvent aussi bonne que *Donnant-donnant*) et *Graduelle* définies dans le texte (F). *Graduelle* arrive en tête : avoir de la mémoire est utile ! D'autres tests ont confirmé que *Graduelle* est supérieure à *Donnant-donnant* : face aux stratégies agressives sans mémoire (*Lunatique, Périodique-cct,* etc.), *Graduelle* tend à avoir le comportement optimal qui est de toujours jouer t, alors que ce n'est pas le cas de *Donnant-donnant*.

A

1. Donnant-donnant : je coopère au premier coup, puis je joue ce qu'a joué mon adversaire au coup précédent.

2. MajoMou : je joue ce que mon adversaire a joué en majorité ; au premier coup ou en cas d'égalité, je coopère.

3. Rancunière : je coopère, mais dès que mon adversaire a trahi, je trahis toujours.

4. Sondeur : aux 3 premiers coups je joue *tcc*; à partir du coup 4, si aux coups 2 et 3 mon adversaire a coopéré, je trahis toujours, sinon je joue *Donnant-donnant*

5. Périodique-*cct* : je joue périodiquement *cct*.

6. Donnant-donnant-dur : je coopère sauf si mon adversaire a trahi l'un des deux coups précédents.

7. Gentille : je coopère toujours.

8. Lunatique : je trahis une fois sur deux au hasard.

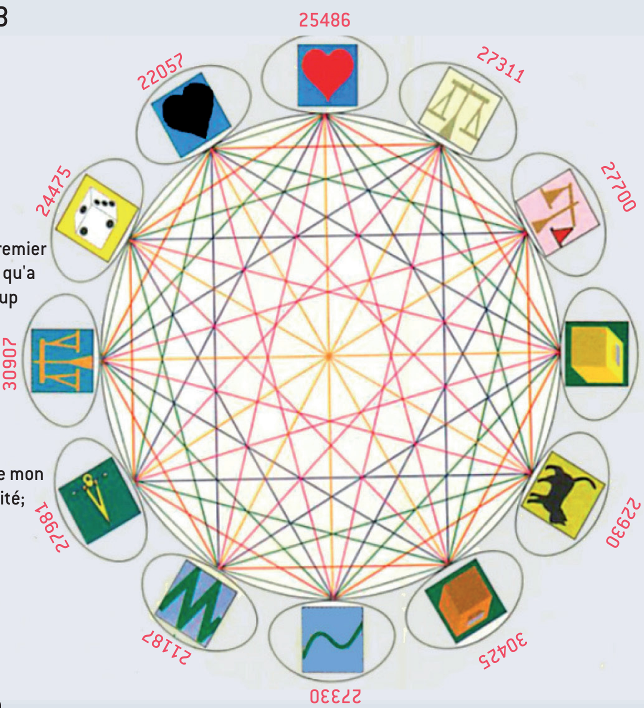
9. Méfiante : je trahis au premier coup, puis après je joue ce qu'a joué mon adversaire au coup précédent.

10. Méchante : je trahis toujours.

11. MajoDur : je joue ce que mon adversaire a joué en majorité; au premier coup ou en cas d'égalité, je trahis.

12. Périodique-ttc : je joue périodiquement *ttc*.

B



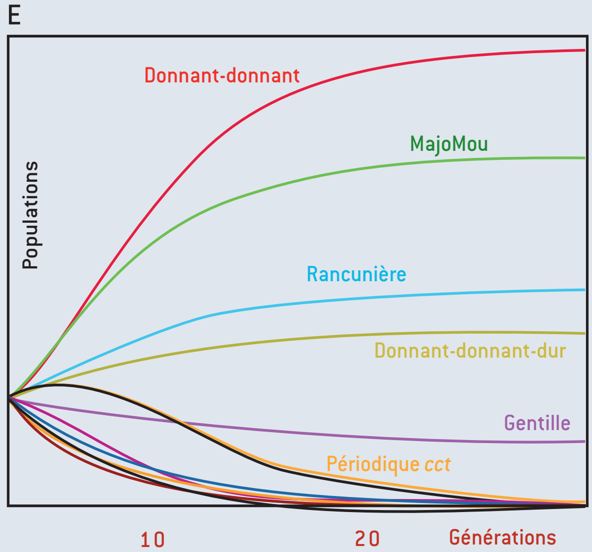
C

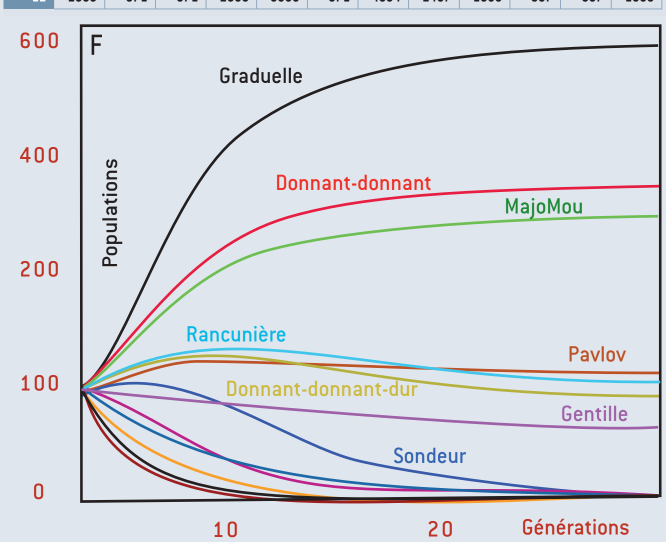
Gain du Joueur 1 / Gain du Joueur 2	Coopère (C)	Trahit (C)
Coopère (C)	3 / 3	5 / 0
Trahit (T)	0 / 5	1 / 1

D

	1	2	3	4	5	6	7	8	9	10	11	12
1	3000	3000	3000	2999	2667	3000	3000	2244	2500	999	2500	1998
2	3000	3000	3000	2999	2001	3000	3000	2095	2500	999	2500	2331
3	3000	3000	3000	1007	3663	3000	3000	2975	1003	999	1003	2331
4	2999	2999	1002	1004	2669	1005	4996	2541	2995	998	2496	1996
5	2667	3666	343	2664	2334	1671	3666	1994	2664	333	3663	1665
6	3000	3000	3000	1010	3331	3000	3000	2634	1003	999	1003	2331
7	3000	3000	3000	6	2001	3000	3000	1486	2997	0	2997	999
8	2248	2645	512	1601	2834	1401	4009	2245	2252	491	2567	1670
9	2500	2500	1003	3000	2669	1003	3002	2254	1000	1000	1000	1999
10	1004	1004	1004	1008	3668	1004	5000	3033	1000	1000	1000	2332
11	2500	2500	1003	2501	2003	1003	3002	2112	1000	1000	1000	2332
12	2003	671	671	2006	3335	671	4334	2497	1999	667	667	1666

E





jouant *t* mais seulement 3 points en jouant *c* ; si l'autre entité trahit, j'obtiens 1 point en jouant *t*, et 0 en jouant *c*.

Il y a dilemme. Collectivement, les deux agents emportent 6 points en jouant [*c*, *c*] alors qu'ils en remportent moins en jouant [*c*, *t*] et encore moins en jouant [*t*, *t*]. L'intérêt collectif est que chacun joue *c*, mais une analyse logique individuelle conduit inévitablement à [*t*, *t*], qui est collectivement le pire cas !

Le dilemme est itéré quand la situation de choix entre *c* et *t* se présente périodiquement aux deux mêmes agents. Jouer consiste alors à choisir une stratégie qui, informée du comportement antérieur de l'adversaire, indique le coup suivant à jouer.

Donnant-donnant, une bonne stratégie

La stratégie *Donnant-donnant* consiste à jouer *c* au premier coup, puis à jouer au coup *n* ce que l'adversaire a joué au coup $n-1$. La stratégie *Périodique-ttc* consiste à jouer *t,* puis *t*, puis *c,* et à recommencer indéfiniment la même séquence des trois coups. Lorsque *Donnant-donnant* rencontre *Périodique-ttc,* la confrontation donne la suite : [*c*, *t*] [*t*, *t*] [*t*, *c*] [*c*, *t*] [*t*, *t*] [*t*, *c*] [*c*, *t*] [*t*, *t*] [*t*, *c*]... Cela rapporte donc en moyenne $(0+1+5)/3=2$ points par coup à *Donnant-donnant* et $(5+1+0)/3=2$ points aussi par coup à *Périodique-ttc.*

Dans certaines stratégies, les coups sont décidés au hasard. La stratégie *Lunatique* joue par exemple *c* dans 50 % des coups et *t* dans 50 % des coups en tirant au hasard avec une pièce de monnaie. Un petit calcul montre que *Lunatique* opposée à *Donnant-donnant* gagne en moyenne $9/4=2{,}25$ points par coup, ce que gagne aussi son adversaire.

Un raisonnement élémentaire montre d'une part que *Donnant-donnant* ne perd jamais plus de 5 points, quel que soit son adversaire et la durée de la rencontre, et d'autre part qu'il ne fait jamais mieux que lui. Dans un tournoi entre un ensemble varié de stratégies (chacune joue contre chaque autre, y compris elle-même, et on totalise les points gagnés), Robert Axelrod a montré que *Donnant-donnant* était une bonne stratégie. Elle ne gagne pas toujours, car certains ensembles de stratégies lui sont défavorables, mais elle se retrouve le plus souvent classée parmi les premières. Elle ne bat jamais personne, mais comme son comportement incite à la coopération, elle obtient de bons scores qui la placent dans le peloton de tête *(voir la figure 1)*.

Deux autres stratégies sont aussi intéressantes que *Donnant-donnant* :

– *Pavlov* : au premier coup, je coopère ; ensuite, si au dernier coup joué, j'ai gagné 3 points ou plus, je rejoue la même chose, sinon je change.

– *Graduelle* : je coopère au premier coup ; ensuite, lorsque mon adversaire me trahit, je le punis au coup suivant (comme *Donnant-donnant* qui répond à une trahison au coup $n-1$ par une trahison au coup *n*), mais je suis plus sévère que *Donnant-donnant,* car je punis mon adversaire en jouant *t* pendant *k* coups consécutifs, où *k* est le nombre de trahisons passées de mon adversaire (mes punitions sont donc « graduelles »). Après une telle phase de rétorsion, je coopère deux fois de suite pour tenter de rétablir la paix.

La stratégie *Graduelle* a été conçue en 1993 par C. Dziengelewski, un lecteur de *Pour la Science* qui participait à un concours organisé par la revue. Le concours a suggéré qu'il n'y a pas de limite à la qualité des stratégies, à condition de les rendre de plus en plus complexes. D'autres études menées depuis, en particulier par Bruno Beaufils du Laboratoire d'informatique fondamentale de Lille conduisent aux conclusions suivantes, généralement acceptées :

– Il n'y a pas de stratégies meilleures que toutes les autres, mais certaines sont mauvaises dans pratiquement tous les environnements possibles, alors que d'autres sont bonnes dans des tournois variés.

– Les bonnes stratégies sont réactives : elles répondent quand on les trahit, prennent le risque de coopérer, commencent par coopérer et, face à un adversaire qui coopère, elles ne trahissent pas, elles savent être indulgentes, après une trahison de l'adversaire elles finissent par pardonner pour renouer la coopération (comme *Graduelle*).

À côté de l'évaluation des stratégies obtenues par des tournois, il existe des méthodes de tests plus subtiles auxquelles ne réussissent que les stratégies très robustes. Les « sélections écologiques » sont l'une de ces méthodes. 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 de chacune des stratégies, ce qui définit une seconde génération. La seconde génération produit selon le même processus une autre génération, etc. Les stratégies gagnantes (celles dont les effectifs sont les plus importants) lors d'une telle évolution sont bonnes dans des arènes variables, leurs bons classements ont donc une signification plus profonde que celui donné par un simple tournoi.

Les résultats obtenus par ces calculs modélisent la sélection naturelle et conduisent à une conclusion étonnante : sauf dans des cas très rares, l'arène finit par n'être occupée que par des stratégies qui ne prennent jamais l'initiative de trahir (c'est le cas de *Donnant-donnant*, de *Graduelle* et de *Pavlov*). Après quelques générations, l'arène est occupée par des stratégies qui ne jouent entre elles que des coups [*c*, *c*], 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, l'évolution élimine toutes les stratégies qui succombent à cette tentation.

Des résultats qui changent la donne ?

L'article de William Press et Freeman Dyson qui a tant étonné les spécialistes en 2012 a un titre provocateur : « Dans le dilemme itéré du prisonnier, il existe des stratégies qui dominent tout adversaire évolutif ». Il a fait sursauter les chercheurs qui pensaient que l'accord sur l'idée d'une convergence vers la coopération généralisée interdisait qu'il puisse exister des stratégies dominantes (donc exploitant les autres), d'autant plus que les stratégies mises en avant par W. Press et F. Dyson ne respectent pas la règle qu'on pensait définitive : il ne faut jamais prendre l'initiative de trahir face

2. Les ZD-stratégies

Une stratégie à mémoire d'un coup est définie par les probabilités p_1, p_2, p_3, p_4 de jouer c lorsque le dernier coup a été respectivement $[c, c]$, $[c, t]$, $[t, c]$ ou $[t, t]$. Notons $Strat(p_1, p_2, p_3, p_4)$ cette stratégie. W. Press et F. Dyson considèrent une classe de stratégies $Strat(p_1, p_2, p_3, p_4)$ dépendant de trois paramètres a, b et c, nommées ZD-stratégies et notées ZD(a, b, c). Les équations générales reliant les paramètres p_1, p_2, p_3, p_4 pour les ZD dans le cas $\{R = 3, S = 0, T = 5, P = 1\}$ sont :

$p_1 - 1 = 3a + 3b + c$;

$p_2 - 1 = 5b + c$;

$p_3 = 5a + c$;

$p_4 = a + b + c$.

W. Press et F. Dyson démontrent que lorsqu'on oppose ZD(a, b, c) à une stratégie *Strat*, et que l'on note G_1 le gain moyen par coup de la première et G_2 le gain moyen par coup de la seconde, alors ces gains moyens vérifient : $a\,G_1 + b\,G_2 + c = 0$. Le gain moyen G_2 ne dépend plus de ce que *Strat* joue ! La stratégie *Strat* est « contrôlée » par la ZD-stratégie.

Lorsque $a = 0$ et $b \neq 0$, on a : $p_1 = 3b + c + 1$, $p_2 = 5b + c + 1$, $p_3 = c$, $p_4 = b + c$ et $G_2 = -c/b$. Autrement dit, *Strat* a un gain moyen indépendant des probabilités qui la définissent, gain qui ne dépend que de la ZD-stratégie qui lui fait face, ZD(a, b, c). Une telle ZD-stratégie est nommée *Égaliseur*.

Prenons un exemple d'*Égaliseur* : $a = 0$; $b = -1/3$; $c = 2/3$; $p_1 = 2/3$; $p_2 = 0$; $p_3 = 2/3$; $p_4 = 1/3$; $G_2 = -c/b = 2$. Voici les résultats des rencontres entre cet *Égaliseur* et diverses stratégies classiques :

Égaliseur = 2,5 ↔ *Donnant-donnant* = 2

Égaliseur = 3 ↔ *Graduelle* = 2

Égaliseur = 3 ↔ *Gentille* = 2

Égaliseur = 1 ↔ *Méchante* = 2

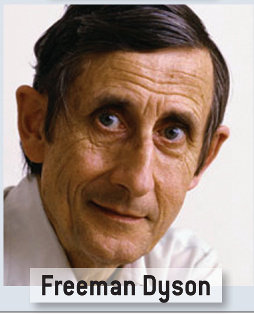
Égaliseur = 2 ↔ *Périodique-cct* = 2

Égaliseur = 1 ↔ *Rancunière* = 2

Égaliseur = 2 ↔ *Égaliseur* = 2

Égaliseur force le gain moyen de l'adversaire, mais cela se fait parfois à ses dépens : contre *Rancunière*, elle n'obtient qu'un point en moyenne par coup et comme *Égaliseur* force les stratégies rencontrées à avoir un faible score, elle en est victime lorsqu'elle joue contre elle-même !

William Press



Freeman Dyson

à un adversaire qui coopère. Pire encore, l'article présentait comme les meilleures possibles des stratégies simples (n'utilisant, pour se déterminer, que le dernier coup joué), ce qui est en contradiction avec l'idée d'un perfectionnement possible illimité des stratégies de plus en plus complexes.

L'étonnement était d'autant plus grand que les résultats avancés étaient des preuves mathématiques dans un domaine où il est difficile de démontrer des résultats, la raison étant que l'espace associé au problème est infini et discret. Il s'agit de l'espace de tous les algorithmes pouvant servir de stratégie, espace qui n'est muni d'aucune structure topologique en lien direct avec les performances des stratégies. Les méthodes habituelles fondées sur le continu ne donnent rien et la richesse énorme de l'espace des stratégies envisageables ne se laisse pas réduire à quelques équations ou considérations logiques. Les résultats de W. Press et F. Dyson remettaient en cause les innombrables conclusions admises et utilisées en théorie de l'évolution, en économie et en sciences sociales.

Le travail de W. Press et F. Dyson s'organise autour de deux arguments tirés de théorèmes. Le premier théorème concerne le dilemme itéré dans une version limitée aux stratégies probabilistes à mémoire d'un coup : les stratégies *Lunatique*, *Donnant-donnant* et *Pavlov* appartiennent à cette catégorie, mais pas la stratégie *Graduelle* qui consulte tous les coups déjà joués.

Stratégies probabilistes à mémoire d'un coup

Une stratégie à mémoire d'un coup est définie par quatre paramètres p_1, p_2, p_3, p_4 indiquant la probabilité de jouer c lorsque le dernier coup a été $[c, c]$, $[c, t]$, $[t, c]$ ou $[t, t]$. Notons $Strat(p_1, p_2, p_3, p_4)$ cette stratégie générale (on supposera que la stratégie coopère au premier coup).

La stratégie *Donnant-donnant* est $Strat(1, 0, 1, 0)$: elle coopère avec une probabilité de 100 % si le dernier coup est $[c, c]$ ou $[t, c]$, et trahit sinon. De même, on vérifie sans peine que la stratégie *Lunatique* est $Strat(1/2, 1/2, 1/2, 1/2)$ et que *Pavlov* est $Strat(1, 0, 0, 1)$.

W. Press et F. Dyson considèrent une classe particulière des stratégies $Strat(p_1, p_2, p_3, p_4)$ dépendant de trois paramètres a, b et c et dénommées ZD-stratégies. Nous les noterons $ZD(a, b, c)$ (*voir la figure 2*). Le *ZD* provient d'un déterminant nul (*zero determinant*) qui apparaît dans les raisonnements conduisant aux théorèmes.

Le plus important est que les deux chercheurs démontrent qu'elles ont la propriété d'avoir un gain moyen lié linéairement à celui de leur adversaire quand celui-ci est pris parmi les stratégies à mémoire d'un coup. Ainsi, quand on oppose $ZD(a, b, c)$ à une stratégie *Strat* et qu'on note G_1 le gain moyen par coup de la première et G_2 le gain moyen par coup de la seconde, ces gains moyens vérifient $aG_1 + bG_2 + c = 0$

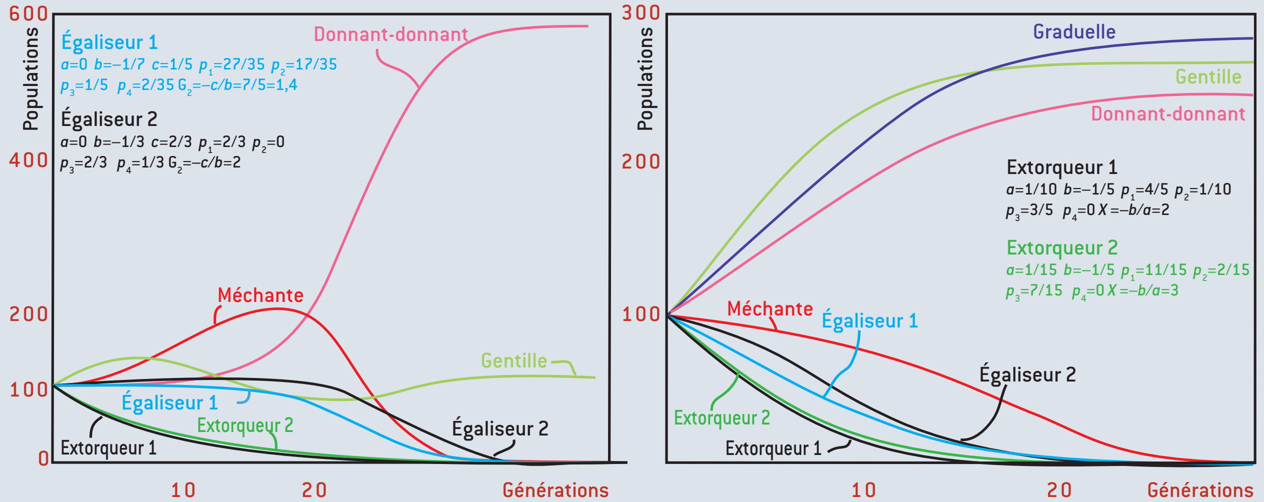
Si $a = 0$ et $b \neq 0$, on a $G_2 = -c/b$: *Strat* a un gain moyen indépendant des probabilités qui la définissent, gain qui ne dépend que de la ZD-stratégie qui lui fait face, $ZD(0, b, c)$! Une telle ZD-stratégie est nommée *Égaliseur*. Contre elle, toutes les stratégies à mémoire d'un coup obtiennent le même gain moyen connu d'avance : $-c/b$. Inutile de se débattre face à une stratégie *Égaliseur*, vous gagnerez $-c/b$ et pas plus ! Les valeurs possibles pour $-c/b$ sont les valeurs situées entre P et R (c'est-à-dire entre 1 et 3 dans le cas classique).

3. Les extorqueurs

Parmi les ZD-stratégies découvertes par W. Press et F. Dyson certaines opèrent une forme d'extorsion. En effet, si $c = -(a + b)P$, on démontre que le gain moyen G_1 de la stratégie ZD contre une autre (laquelle obtient le gain moyen de G_2) vérifie : $G_1 - P = X(G_2 - P)$ avec $X = -b/a$. En clair, si la seconde veut gagner plus, donc augmenter $(G_2 - P)$, cela entraîne mécaniquement que la stratégie ZD augmente son gain moyen, dont l'écart à P est toujours X fois l'écart à P du gain de la seconde.

Les quatre paramètres définissant ces stratégies *Extorqueurs* sont donnés par les équations : $p_1 = 2a + 2b + 1$; $p_2 = 4b - a + 1$; $p_3 = 4a - b$; $p_4 = 0$.

Prenons deux stratégies *Extorqueurs* de cette catégorie, deux stratégies *Égaliseurs* dont les paramètres sont donnés ci-dessous et regardons leurs évolutions avec les stratégies classiques, dans un cas sans *Graduelle* et dans l'autre cas avec. Un des gros défauts des stratégies *Extorqueurs* est qu'elles jouent mal contre elles-mêmes et contre n'importe quelle autre stratégie *Extorqueur*. Dans la simulation évolutive, les *Égaliseurs* et les *Extorqueurs* ne subsistent pas et *Graduelle* prend la première place.



La stratégie *ZD*(0, −1/3, 2/3) *(voir la figure 2)* qui est *Strat*(2/3, 0, 2/3, 1/3) force le gain moyen de son adversaire, qui sera de 2 points en moyenne par coup, quelle que soit la stratégie à mémoire d'un coup prise par l'adversaire. Il y a mieux encore : les ZD-stratégies dénommées *X-Extorqueur*. Elles ne vous permettent d'avoir un bon résultat que si vous leur en accordez un proportionnellement meilleur. Ce sont celles où $c = -(a + b)P$. En jouant contre elle, la relation entre votre gain moyen G_2 et le sien G_1 est $G_1 - P = X(G_2 - P)$, où $X = -b/a$.

Cela signifie que ce qu'elles gagnent en moyenne en plus de P (pris égal à 1 dans le cas usuel) est proportionnel à ce que leur adversaire gagne en moyenne en plus de P. Si vous jouez contre un *X-Extorqueur*, vous ne pouvez améliorer votre score moyen qu'en améliorant le sien : il vous taxe en proportion X de vos revenus !

Considérons *ZD*(1/10, −1/5, 1/10) (*Extorqueur 1* de la figure 3) qui est la stratégie *Strat*(4/5, 1/10, 3/5, 0). C'est une stratégie *2-Extorqueur* : elle double pour elle l'écart à P. Par exemple, quand elle joue contre *Pavlov*, ce dernier gagne en moyenne 1,62 (0,62 point en plus de 1) alors que l'*Extorqueur* obtient en moyenne 2,24 (1,24 point en plus de 1).

Pas plus de mémoire

Le second théorème important de l'article de W. Press et F. Dyson indique que si, dans une partie supposée infinie, la stratégie A est face à une stratégie B ayant une mémoire de k coups, il existe alors une stratégie A' qui obtient le même score moyen face à B et qui n'a qu'une mémoire de k coups. Pour obtenir un certain résultat face à une stratégie donnée, il n'est jamais nécessaire d'avoir plus de mémoire qu'elle.

La combinaison des deux résultats mathématiques de W. Press et F. Dyson affirme que face à une stratégie *Égaliseur* ou *Extorqueur*, ce ne sont pas seulement toutes les stratégies à mémoire d'un coup qui se trouvent contraintes, mais toutes les stratégies à mémoire finie. On est tenté d'en conclure que : « les stratégies mémorisant plus que le dernier coup sont inutiles et que l'on dispose avec les ZD-stratégies de stratégies dominantes dans l'absolu pour le dilemme itéré du prisonnier ».

Certains ont d'ailleurs interprété ainsi les théorèmes démontrés, et le titre retenu pour leur article suggère que c'est le cas de W. Press et F. Dyson (bien que le contenu de leur article soit plus prudent). Pourtant, la double affirmation sur l'inutilité des stratégies à mémoire étendue et sur la dominance absolue des ZD-stratégies est totalement fausse, comme l'ont prouvé plusieurs articles parus en 2013 *(voir la bibliographie)*. Voici pourquoi.

Considérons d'abord l'affirmation de dominance des ZD-stratégies. On sait depuis longtemps qu'au dilemme itéré du prisonnier, battre son adversaire (obtenir plus de points que lui) peut se faire aux dépens de celui qui gagne et qui aurait pu obtenir plus de points en moyenne en acceptant d'être battu. La stratégie *Méchante* (qui joue toujours *t*) bat toute autre stratégie (c'est une évidence) ; par exemple, sur une partie de 100 coups contre *Donnant-donnant*, elle obtient 104 points, alors que *Donnant-donnant* en gagne 99. La stratégie *Gentille* (qui joue toujours *c*) ne bat pas *Donnant-donnant* et gagne 300 points en 100 coups contre *Donnant-donnant*, laquelle gagne aussi 300 points. Contre *Donnant-donnant*, *Méchante* gagne peut-être, mais elle a tort de gagner car en faisant comme *Gentille*, elle obtiendrait un bien meilleur score.

Mieux vaut s'entendre avec son adversaire

La plupart des ZD-stratégies sont un peu dans la même situation : elles ne battent leur adversaire qu'en renonçant elles-mêmes à de bons scores. Les *Extorqueurs* gagnent sur les stratégies auxquelles on les oppose, mais aux dépens du total des points gagnés. D'ailleurs, une stratégie *X-Extorqueur* (avec $X > 1$) qui joue contre elle-même n'obtient (d'après la théorie qu'on vérifie par simulation) qu'un seul point en moyenne par coup, ce qui est très médiocre. Il n'est pas vrai qu'être une bonne stratégie, c'est battre son adversaire. Mieux vaut ne pas le battre, bien s'entendre avec lui et marquer beaucoup de points.

Le cas de *Donnant-donnant* est remarquable et tel qu'on a l'impression d'un paradoxe quand on énonce ses propriétés : *Donnant-donnant* ne bat aucune stratégie individuellement et se fait battre par de nombreuses stratégies et pourtant, c'est une bonne stratégie qui gagne de nombreux tournois et compétitions évolutives. Elle gagne non pas parce qu'elle force les autres à gagner moins qu'elle, comme le fait une stratégie *Extorqueur*, mais parce qu'elle punit les stratégies qui ne veulent pas coopérer. Elle force la coopération. Face à elle, ou bien vous gagnerez peu de points, ou bien vous coopérerez, ce qui sera bon pour elle et pour vous.

C'est donc un contresens que croire que les ZD-stratégies sont bonnes. Dans des rencontres un contre un, elles dominent, comme le fait *Méchante*, mais globalement elles jouent piteusement ! Outre l'erreur de croire que battre son adversaire, c'est gagner des points, un autre oubli a conduit à croire que les ZD-stratégies étaient dominantes : pour s'imposer, il faut jouer correctement contre soi-même. C'est important dans les tournois, mais plus encore dans les compétitions évolutives. En effet, si vous l'emportez lors des premières générations, l'arène se peuple de stratégies identiques à vous, et vous allez donc très fréquemment les rencontrer. Si vous jouez mal face à vous-même, cela se retourne contre vous.

Rien n'est faux dans les résultats mathématiques de W. Press et F. Dyson, mais en n'abordant que le problème « qui gagne dans un combat un contre un ? » et en oubliant les questions « Combien de points sont gagnés ? » et « Face à toi-même, te fais-tu du tort ? », les théorèmes démontrés ne permettent pas de conclure que les ZD-stratégies sont de bonnes stratégies.

Le résultat de W. Press et F. Dyson sur l'inutilité pour une stratégie d'avoir de la mémoire est juste dans le sens précis de ce qu'ils démontrent : si A est face à une stratégie B ayant une mémoire de *k* coups, il existe une stratégie A' qui obtient le même score moyen face à B et qui n'a qu'une mémoire de *k* coups. Mais cela ne signifie pas que face à deux stratégies B et C ayant une mémoire de *k* coups, il existe une stratégie A' à mémoire de *k* coups qui fasse le même score face à B et face à C. En effet, celle, A', qui peut remplacer A face à B n'est pas nécessairement la même que celle, A", qui peut remplacer A face à C. Le résultat de W. Press et F. Dyson sur l'inutilité de la mémoire est valable uniquement dans les combats un contre un ; il cesse d'être vrai dès qu'on envisage des tournois ou des évolutions.

Il est très utile que des chercheurs viennent explorer un domaine qui n'est pas le leur, et si W. Press et F. Dyson n'avaient pas été aussi confiants en leurs résultats, ils n'auraient pas proposé un titre provocateur à leur article – titre qui, après analyse, est faux – et ils n'auraient pas attiré l'attention sur ce qui, malgré tout, présente un véritable intérêt. *Se non è vero è bene trovato !* ■

L'AUTEUR

J.-P. DELAHAYE est professeur à l'Université de Lille et chercheur au Laboratoire d'informatique fondamentale de Lille (LIFL).

BIBLIOGRAPHIE

C. Adami et A. Hintze, **Evolutionary instability of zero determinant strategies demonstrates winning is not everything**, *Nature Comm.*, vol. 4, 2193, 2013 (en accès libre).

C. Hilbe *et al.*, **Evolution of extorsion on iterated prisoner's dilemma games**, *PNAS*, vol. 110, pp. 6913-6918, 2013.

• W. Press et F. Dyson, **Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent**, *PNAS*, vol. 109, pp.10409-10413, 2012.

K. Sigmund, **The Calculus of Selfishness**, Princeton University Press, 2010.

B. Beaufils, **Modèles et simulations informatiques des problèmes de coopération entre agents**, Thèse de doctorat, Université de Lille, 2000.

M. Boerlijst *et al.*, **Equal pay for all prisoners**, *American Mathematical Monthly*, vol. 104, pp. 303-307, 1997.

W. Poundstone, **Le dilemme du prisonnier**, Cassini, 1993.

B. Beaufils, J.-P. Delahaye et P. Mathieu, **Our meeting with Gradual : A good strategy for the iterated prisoner's dilemma**, *Proceedings of Artificial Life V*, MIT Press/ Bradford Books, pp. 202-209, 1996.