Ahmed Amine Bennani et Souhail Cadi
Ecole Polytechnique
ahmed.bennani.x17@polytechnique.edu
souhail.cadi@polytechnique.edu

ÉCOLE POLYTECHNIQUE

IP PARIS

DILEMME DU PRISONNIER

CONTENTS

1 MISE EN CONTEXTE

1.1 Dilemme du prisonnier classique

Le dilemme du prisonnier est l'exemple le plus connu de la Théorie des jeux, théorie qui cherche à rendre compte des stratégies et des décisions des individus rationnels en interaction dans le cadre imparti par des règles du jeu. Le comportement de chaque joueur est analysé en fonction des comportements observés ou attendus des autres joueurs. Le jeu du dilemme du prisonnier a été proposé pour la première fois par Merrill Flood en 1951. Il a été formalisé et définie par Albert W. Tucker. Le nom fait référence à la situation hypothétique suivante.

Tucker suppose deux prisonniers (complices d'un crime) retenus dans des cellules séparées et qui ne peuvent pas communiquer. Les choix qui s'offrent aux prisonniers sont :

- si un seul des deux prisonniers dénonce l'autre, il est remis en liberté alors que le second obtient la peine maximale (10 ans) ;
- si les deux se dénoncent entre eux, ils seront condamnés à une peine plus légère (5 ans) ;
- si les deux refusent de dénoncer, la peine sera minimale (6 mois), faute d'éléments au dossier.

Ce "jeu" peut être généralisé à n'importe quelle situation où deux joueurs se trouvent dans une situation de non-coopération où la meilleure situation globale est que les deux coopèrent, mais le pire résultat individuel est d'être un joueur coopérant tandis que l'autre joueur fait défaut. L'objectif de chaque prisonnier est de minimiser sa peine. La solution optimale est la coopération entre les deux prisonniers mais le manque d'information et l'impossibilité de communication crée une incertitude sur le comportement du complice. La théorie des jeux implique que l'équilibre de Nash correspond à la situation ou chacun optimise son choix par rapport à l'autre en choisissant de dénoncer.

Néanmoins, il est évident qu'il n'y a pas de **solution optimale unique** pour un prisonnier, et c'est pour cela que c'est un dilemme. Ce modèle très simple de la théorie des jeux semble appréhender en miniature les tensions entre cupidité individuelle et intérêts de la coopération collective. Pour cette raison, il est devenu un des modèles les plus utilisés en sociologie, biologie et économie.

1.2 Solution théorique pour le dilemme en 1 tour

Pour résoudre le dilemme dans un jeu non répété (c'est-à-dire, joué une seule fois), on utilise la notion de **stratégie dominante**. Une stratégie est dominante lorsqu'elle va correspondre au choix que j'ai intérêt à faire quelle que soit la stratégie adoptée par l'autre joueur.

En effet, si le joueur adverse coopère et que je le dénonce, je serais remis en liberté alors que si je coopère, ma peine serait de 6 mois. Si le joueur adverse me dénonce et que je le dénonce également ma peine sera de 5 ans alors que si je coopère, ma peine sera de 10 ans. Comme on peut le voir dans les deux cas de figure, en considérant uniquement mon bien être personnel, le meilleur choix est de dénoncer.

La solution d'équilibre en stratégies dominantes, qui est aussi solution de l'équilibre de Nash, est donc non coopérative. La solution est considérée comme non optimale pour la société (ici representée par la somme des deux joueurs). La solution optimale pour la société est celle qui minimise la peine pour les deux joueurs, c'est-à-dire la stratégie où les deux joueurs coopérent.

Le dilemme du prisonnier incarne l'idée fondamentale selon laquelle la confrontation des intérêts individuels ne débouche pas nécessairement sur l'optimum collectif.

1.3 Le dilemme répété

Dans son livre *The Evolution of Cooperation* publié en 1984. Robert Axelrod étudie une extension classique de ce dilemme : le jeu se répète, et les participants gardent en mémoire les précédentes rencontres. En effet, dès 1980, il a decidé d'organiser un tournoi de différentes stratégies pour le dilemme du prisonnier. Il a invité un certain nombre de théoriciens du jeu bien connus à soumettre des stratégies à exécuter par ordinateur. Dans le tournoi, les programmes ont joué des parties les uns contre les autres et contre eux-mêmes à plusieurs reprises. Chaque stratégie spécifiait si elle devait coopérer ou non en se basant sur les mouvements précédents de son adversaire.

Quand on répète ce jeu durablement dans une population, les joueurs qui adoptent une stratégie agressive y perdent au long terme, alors que les joueurs apparemment plus coopératifs voient leur « altruisme » finalement récompensé : le dilemme du prisonnier n'est donc plus à proprement parler un dilemme. Axelrod y a vu une explication de l'apparition d'un comportement altruiste dans un contexte d'évolution darwinienne par sélection naturelle.

Comme souligné auparavant, la répétition du jeu peut inciter l'un des joueurs à « signaler » à l'autre qu'il est prêt à coopérer au premier tour pour l'inciter à faire de même aux tours suivants. On peut alors envisager que la solution théorique soit celle de la coopération et non celle de la défection. Mais, il faut cependant distinguer le cas dans lequel le dilemme est répété un nombre fini de fois, de celui dans lequel il est répété un nombre infini de fois.

1.4 Solution théorique pour le dilemme répété

Si le nombre N d'itérations est fini et connu, l'équilibre de Nash est de systématiquement dénoncer son complice, comme pour N=1. Cela se montre simplement par récurrence descendante ou induction à rebours:

- au dernier coup, sans sanction possible de la part de l'adversaire, on a intérêt à trahir ;
- ce faisant, à l'avant-dernier coup, comme on anticipe que l'adversaire trahira quoi qu'il arrive au coup suivant, il vaut mieux trahir aussi ;
- on poursuit le raisonnement jusqu'à refuser de coopérer à tous les coups.

Toutefois si l'on suppose que le nombre d'itérations est infini ou aléatoire, il n'y a pas de solution théorique d'équilibre en stratégie dominante unique. En effet, lorsque les joueurs ne connaissent pas la durée pendant laquelle le jeu est répété, ils peuvent être incité à adopter une stratégie de réputation de coopération. En coopérant, le joueur peut connaitre des pertes importantes à court terme mais peut espérer des gains futurs à long terme lorsque les deux joueurs coopèrent.

Dès lors, il y a un compromis à faire entre la préférence pour le court terme ou le long terme. Plusieurs stratégies possibles vont pouvoir émerger, puisqu'avec un horizon infini les joueurs vont essayer de s'adapter à l'adversaire et ceci à travers différents manières.

Dans la suite de notre rapport, nous allons nous intéresser à la simulation du tournoi d'Axelrod avec un nombre N fini de tours mais inconnus des joueurs, avec plusieurs stratégies différentes. Puis nous allons observer les stratégies qui auront le meilleur résultat en faisant variant un certain nombre de paramètres.

2 SIMULATION DU TOURNOI D'AXELROD

2.1 Présentation générale de la simulation:

Nous avons simulé le tournoi d'Axelrod qui correspond au jeu du dilemme du prisonnier répété où les joueurs ne connaissent pas le nombre de fois qu'ils jouent contre leur adversaire. Globalement, la simulation se fait en 5 grandes étapes :

1. on a crée une population de taille P=2N constituée de plusieurs types d'individus aux stratégies différents
2. Chaque jour, on choisit aléatoirement N couples qui vont faire des D duels pendant la journée
3. On cumule les scores des duels quotidiens avec les scores totaux de chaque individu
4. On supprime la mémoire de chaque joueur pour qu'il se comporte à chaque jour de la même façon
5. On répète cette procédure pendant J jours

2.2 Les différents types de stratégie:

Nous réalisons notre simulation du tournoi d'Axelrod classique avec 6 types de stratégies :

- **Agressif** : cette stratégie consiste à dénoncer son complice dans toutes les situations
- **Coopératif** : cette stratégie consiste à coopérer avec son complice dans toutes les situations
- **Aléatoire** : cette stratégie consiste à dénoncer ou coopérer de manière aléatoire indépendemment du passé
- **Tit For Tat** : cette stratégie consiste à commencer de manière coopérative au moins les 2 premiers tours et de dénoncer uniquement s'il s'est fait dénoncé lors du dernier tour.
- **Tit For Two Tats** : cette stratégie consiste à commencer de manière coopérative au moins les 2 premiers tours et de dénoncer uniquement s'il s'est fait dénoncé les 2 derniers tours.
- **Two Tits For Tat** : cette stratégie consiste à commencer de manière coopérative au moins les 2 premiers tours et de dénoncer 2 fois de suite dans le cas où il s'est fait dénoncé le tour préalable.

Pour que ces stratégies puissent être appliquées , il est nécessaire que les joueurs enregistrent le résultat des deux derniers duels pour pouvoir réagir en conséquent. Dès lors, on représente les décisions de chaque stratégie par une matrice 2x2 qui permet de donner le comportement dans chacun des 4 cas de figure issues des 2 derniers coups.

De ce fait, nous définissons des matrices de stratégies de la manière suivante.

- la colonne 0 correspond à la réaction dans le cas où le dernier round l'adverseaire a été coopératif.
- La ligne 0 de cette colonne 0 correspond à la réaction dans le cas où l'adversaire a été coopératif les 2 derniers tours.
- La ligne 1 de cette colonne à correspond à la réaction dans le cas où l'adversaire a été coopératif lors du dernier tour mais agressif lors de l'avant dernier tour.

Listing 1. Les matrices de stratégie

```
class strategy():

    def __init__(self,coup1,coup2,mat):
        self.coup1=coup1
        self.coup2=coup2
        self.prob_mat=mat


    def __eq__(self,strategy):
        return(self.prob_mat==strategy.prob_mat)


    Agression = strategy(0,1,np.ones((2,2)))
    Cooperation = strategy(0,0,np.zeros((2,2)))
    Alea = strategy(0,0,0.5*np.ones((2,2)))
    Tit_For_Tat = strategy(0,0,np.array([[0,1],[0,1]]))
    Tit_For_Two_Tats= strategy(0,0,np.array([[0,0],[0,1]]))
    Two_Tits_For_Tat = strategy(0,0,np.array([[0,1],[1,1]]))
```

Ci dessus, on retrouve les matrices de stratégie de base que nous avons utilisé lors de cette simulation. Par convention, on décide que lors du coup 1, il n'y a pas d'agression. La stratégie Aléatoire ne sera pas prise en compte dans notre simulation du Tournoi d'Axelrod car nous voulions conserver un codage binaire. Nous reviendrons plus tard sur cela.

2.3 Objectif de la simulation

L'objectif de la simulation est de calculer la **distribution de score des joueurs de chaque stratégie** pour déterminer celle qui est la plus efficace dans le cadre du jeu du dilemme du prisonnier.

Nous avons choisi de définir une matrice de d'années de condamnations qui permet d'octroyer les années de condamnation dans chacune des situations de la façon suivante :

		Joueur 1	
		0	1
Joueur 2	0	1 et 1	5 et 0
	1	0 et 5	3 et 3

Figure 1. Matrice d'années de condamnation

Dès lors à chaque round de confrontation, on cumule le nombre d'années de condamnation de chaque prisonnier comme on peut le voir à travers le bout de code suivant.

Listing 2. Cumul du nombre d'années de condamnation à chaque round

```
def Round(self):
        self.round_current+=1
        #1st prisoner's choice
        coup1 = self.p1.coup(self.p2)
        #2nd prisoner's choice
        coup2 = self.p2.coup(self.p1)
        ag=0
        if coup1==0 and coup2==0:
            years1=1
            years2=1 #
        if coup1==1 and coup2==1:
            years1=3
            years2=3
            ag=2
        if coup1==0 and coup2==1:
            years1=5
            years2=0
            ag=1
        if coup1==1 and coup2==0:
            years1=0
            years2=5
            ag=1
        self.p1.update(coup1,years1)
        self.p2.update(coup2,years2)
        return(ag)
```

Ensuite, après avoir simuler l'ensemble des duels, on comptabilise les scores de chaque individu de chaque stratégie pour pouvoir les représenter sur un graphe, comme on pourra le voir dans la partie suivante.

3 ANALYSE DE LA SIMULATION DU TOURNOI

3.1 Une simulation de référence

Nous allons étudier l'évolution des résultats pour chaque stratégie en faisant varier plusieurs paramètres :

- Le nombre total de prisonniers, dont la valeur conventionnelle sera : **250**
- Le proportion relative des stratégies qui sera conventionnellement **uniforme**
- Le nombre de confrontations par jour pour chaque couple selectionné aléatoirement dont la valeur conventionnelle sera : **20**
- Le nombre total des journées de confrontations dont la valeur conventionnelle sera : **300**

Le graphe suivant présente les résultats des scores de chaque individu pour chaque stratégie dans le cas conventionnel décrit préalablement.

Figure 2. Simulation des scores pour le cas standard

Comme le score correspond au nombre d'années de condamnation, **plus le score est élevé moins la stratégie est efficace**.

On peut voir que globalement :

- les stratégies punitives TFT, TFTT et TTFT sont les **plus efficaces**, et comptabilisent à peu près le même nombre d'années de condamnation.
- la stratégie coopérative est moins efficace que les punitives mais meilleure que la stratégie agressive
- la stratégie agressive est celle qui cumule le plus d'années de condamnation

La stratégie coopérative consiste à coopérer quelle que soit la situation. C'est une très bonne stratégie lorsque l'adversaire est coopératif ou punitif puisqu'à chaque round les 2 décident de coopérer ce qui minimise la peine à 1 an. Toutefois, face à un adversaire agressif, le Coopératif n'optimise pas son résultat et prend toujours une condamnation maximale de 5 ans.

La stratégie agressive aura une peine nulle uniquement contre les adversaires coopératifs et dans tous les autres cas de figures, elle aura globalement une condamnation de 3 ans à chaque confrontation. Ainsi, son comportement ne profite pas de la possibilité d'avoir des peines légères en coopérant avec les adversaires qui sont enclin à le faire.

Les stratégies punitives Tit for Tat (TFT), Two Tits for Tat(TTFT) et Tit for Two Tats (TFTT) vont elles de leur coté choisir de coopérer en priorité et tant qu'il n'y a pas d'agression, elles continuent leur comportement coopératif ce qui leur permet de minimiser leur peine avec des adversaire punitifs ou coopératifs. Face à un Agressif par contre qui multiplie les attaques, les punitifs vont subir au tout début puis vont aussi commencer à dénoncer jusqu'à la fin de la journée. Ils optimisent ainsi leur peine puisque leur comportement punitif minimise leur peine face à un agressif. Ce type de stratégies **qui s'adaptent à l'adversaire** est ainsi la plus efficace dans ce format du tournoi d'Axelrod.

On remarquera que dans cette situation aux comportements déterministes, les TFT et TTFT ont systématiquement **les mêmes comportements** et cumulent pour un même type de confrontation le même score, et sont les stratégies les plus efficaces. Contre les Punitifs et Coopératifs, ils coopérent tout au long du duel, et contre les Agressifs ils contre attaquent dès le 3ème tour. La stratégie TFTT a globalement le même comportement mais décalé à un tour près dans leurs confrontations face aux Agressifs, puiqu'ils se mettent à contre attaquer qu'à partir du 4ème tour, après avoir reçu deux agressions successives, ce qui explique le fait qu'ils ont un score légèrement supérieur aux autres punitifs. Toutefois, ces stratégies punitives auront des comportements un peu plus différent et auront **chacune leur importance** lors de la dernière partie lorsqu'on simulera le jeu sur plusieurs générations.

3.2 Observation de l'impact du nombre totale de prisonniers

Dans cette partie, on simule le tournoi avec les valeurs prises par convention et en faisant varier le population P.

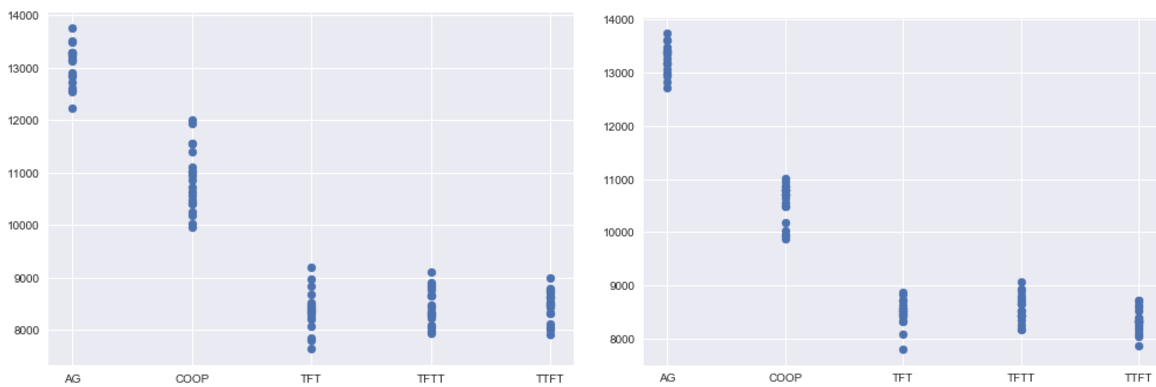

Figure 3. Comparaison avec le cas où on réduit la population

La figure 3 permet de faire une comparaison entre le cas conventionnel (graphe à gauche) et le cas où on diminue la population à 100 individus (graphe de droite). On remarque que globalement, il n'y a pas de grand changement dans la courbe et qu'on peut déduire les mêmes résultats que pour le cas à 300 individus.

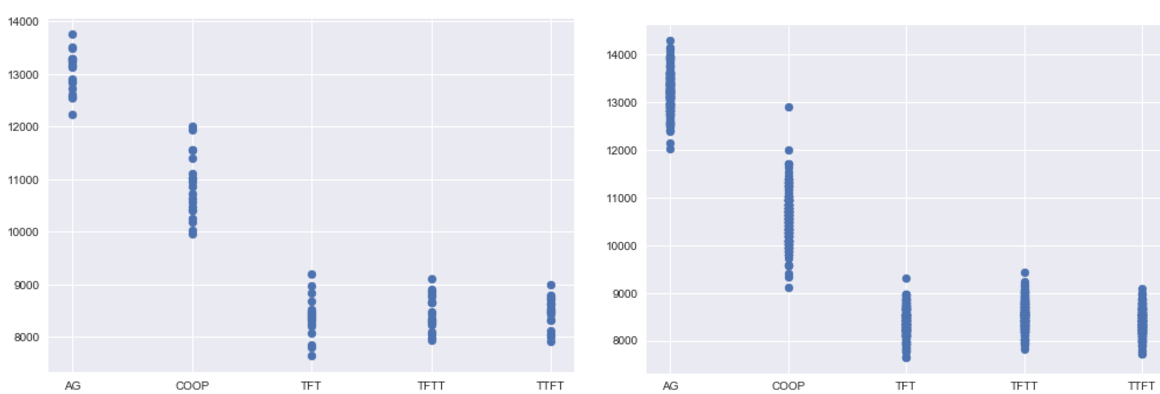

Figure 4. Comparaison avec le cas où on augmente la population

Le graphe de droite correspond au cas où nous augmentons la population à 1000 individus. On remarque que globalement, il n'y a pas de grand changement dans la courbe et qu'on peut déduire les mêmes résultats que pour le cas à 300 individus. Toutefois, on observe une **distribution plus large** pour les coopératifs. Ceci s'explique par le fait que le grand nombre de prisonniers, il y'a de plus grosses différences entre la répartition des types d'adversaires. Les coopératifs qui ont de la chance de jouer contre moins d'agressifs ont un score de condamnation inférieur à ceux qui ont moins de chance.

3.3 Observation de l'impact des proportions des stratégies

Nous allons maintenant observer l'influence des proportions des stratégies sur les scores finaux des individus de chaque stratégie.

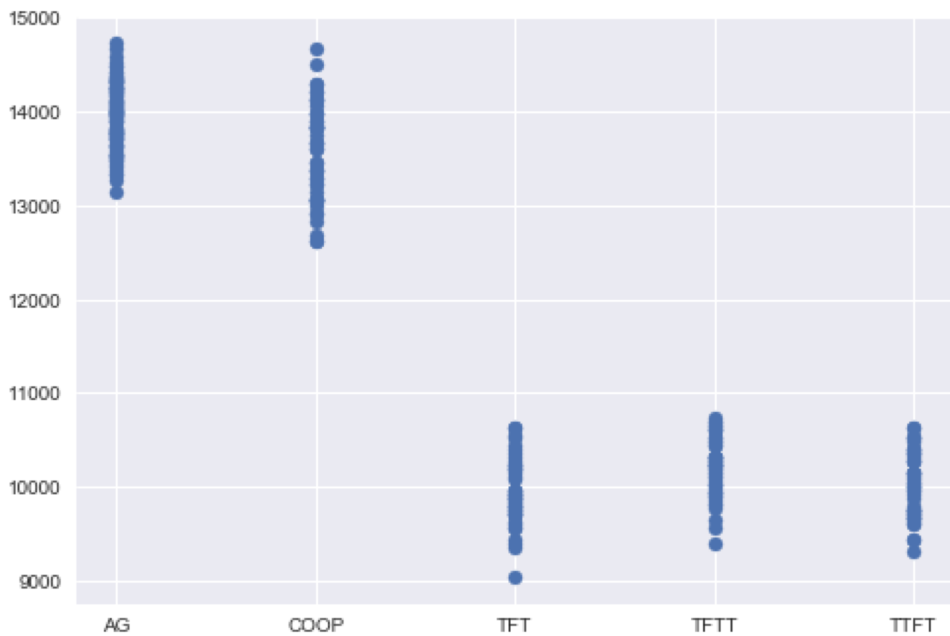

Figure 5. Simulation avec 2 fois plus d'agressifs

Le graphe ci dessus représente les scores dans le cas où on augmente la proportion de joueurs agressifs à 1/3 avec un nombre totale de 300 joueurs.

On remarque que globalement le graphe ne change pas par rapport au cas d'une répartition uniforme, sauf pour la catégories des coopératifs. Ceux ci font face à un **plus grand nombre d'Agressifs** contre qui ils cumulent les condamnations maximales ce qui déplace leur score vers le haut.

Parallèlement, on remarque que les joueurs de stratégies punitives conservent globalement leur classement relatif et ont été légèrement décalé vers le haut, ce qui s'explique par le fait que leurs confrontations avec les nouveaux agressifs n'influent pas beaucoup leur score car ils se neutralisent en se dénonçant.

Les Agressifs eux sont légèrement translatés vers des scores plus élevés de la même façon que les Punitifs, car ils font face à plus d'Agressifs et donc moins de Coopératifs contre qui ils profitaient en prenant des peines nulles.

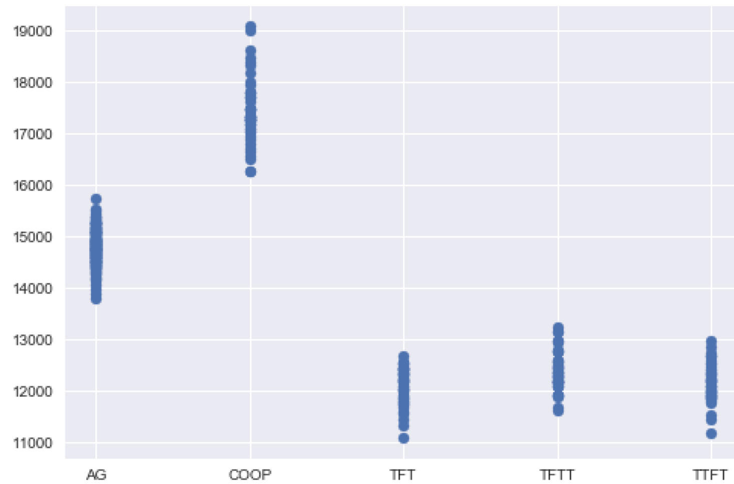

Figure 6. Simulation avec une population de 400 à moitié agressive

Lors de la simulation ci-dessus, le phénomène est encore plus marqué avec les coopératifs qui se retrouvent avec la pire stratégie de jeu.

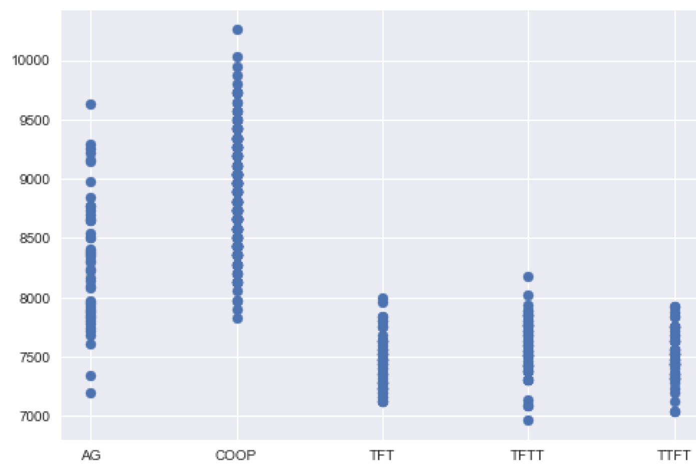

Figure 7. Simulation avec une population de 400 à moitié coopérative

De même, on remarque que lorsque le nombre de coopératifs devient relativement grand comme c'est le cas sur le graphe ci-dessus, l'ensemble des scores diminue globalement, mais les Coopératifs se retrouvent avec le score le plus élevé relativement, ce qui signifie encore une fois que c'est la stratégie la moins efficace.

Cela s'explique par le fait que globalement, il y a moins de condamnations puisque les Coopératifs permettent aux autres de minimiser leur peine. Toutefois, les Coopératifs vont être fortement sanctionnés par les Agressifs qui vont profiter de leur clémence et éviter un maximum de peine, au point où ils se retrouvent avec un score globalement inférieur.

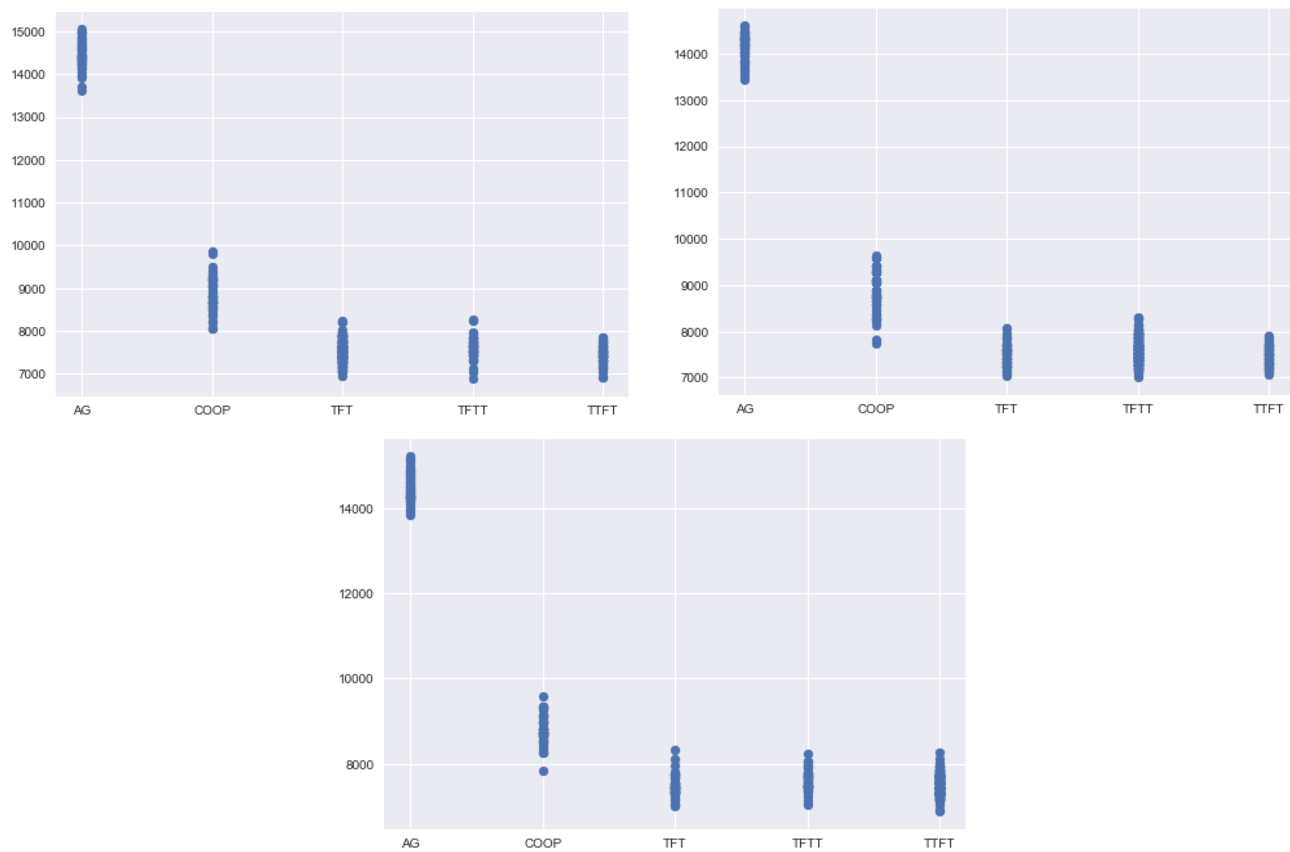

Figure 8. Comparaison avec le cas où on augmente la proportion de Punitifs

Dans les 3 graphes ci-dessus, nous avons augmenté la proportion des punitifs jusqu'à 1/2, que ce soit les Tit For Tats (en haut à gauche), les Tit For Two Tats (en haut à droite) et Two Tits for Tats (en bas). Dans les 3 cas, on observe que les répartitions sont à peu près identiques et renforcent le résultat dans le cas d'équirépartition.

En effet, il y a un effet de répulsion des agressifs vers des condamnations supérieurs car ceux ci prennent beaucoup de peine lors de leurs duels avec les punitifs (3 ans à partir de la 3ème confrontation de la journée). Tandis que de l'autre côté, les autres punitifs et coopératifs réussissent à minimiser leur peine (1 an par confrontation) et ont alors un résultat relatif meilleur et ont donc des stratégies plus efficaces.

Globalement, on observe que les stratégies punitives restent les plus efficaces encore une fois, ce qui renforce les résultats préalables.

3.4 Observation de l'impact du nombre de jours de confrontations

Dans le cas standard, nous avons choisi de réaliser le tournoi en 300 jours de duels. Nous allons observer l'effet du nombre de jours J sur le graphe des scores par stratégie.

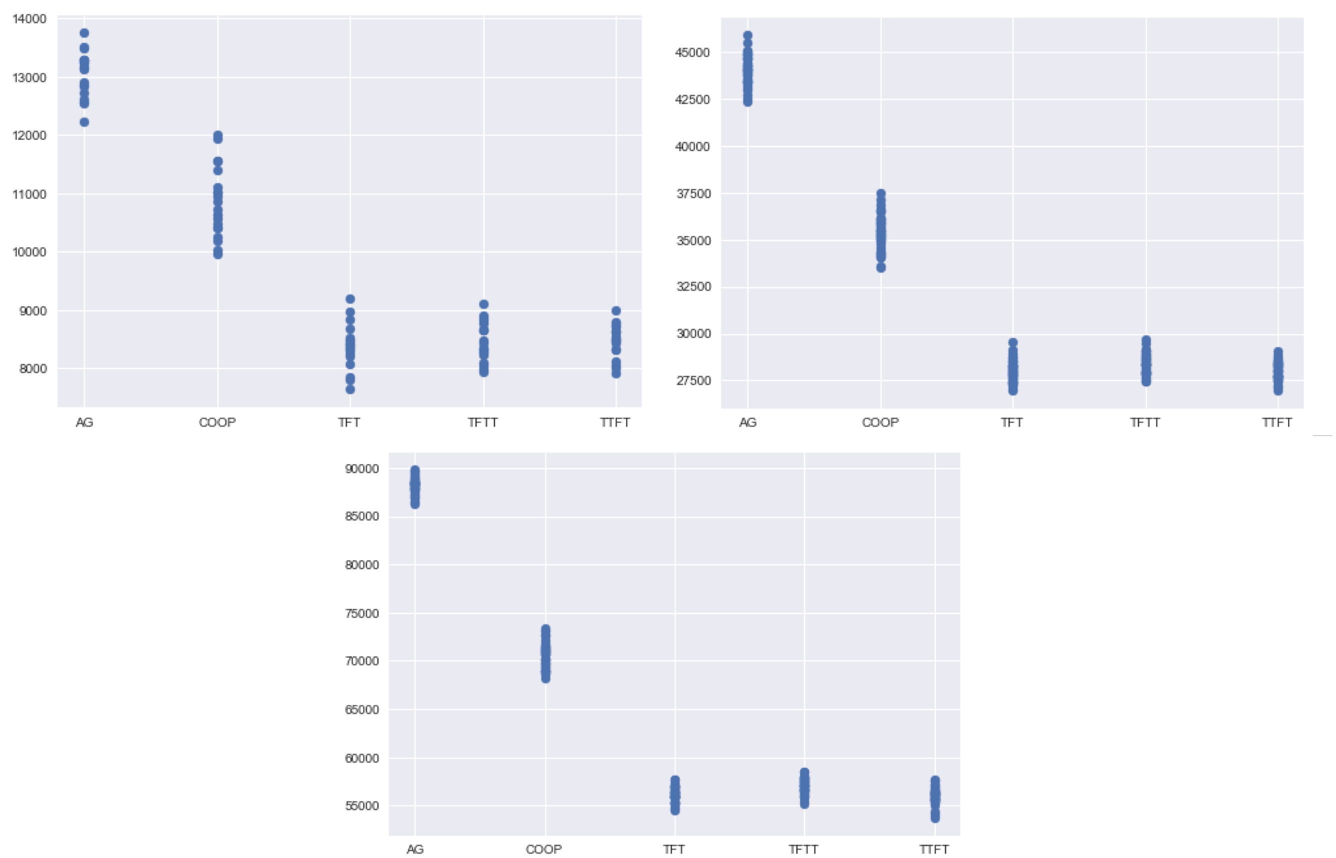

Figure 9. Comparaison avec l'augmentation du nombre de jours de duels

Les graphes ci-dessus montre l'évolution des scores avec l'augmentation du nombre de jours. Le graphe en haut à gauche correspond au cas standard décrit préalablement c'est à dire 300 jours , le graphe en haut à droite correspond au cas 1000 jours et celui du bas correspond au cas de 2000 jours.

On remarque que lorsqu'on augmente le nombre de jours, le **classement des stratégies** reste globalement **le même** avec les Agressifs qui ont la stratégie la moins efficace, les Coopératifs qui se trouvent au milieu et les Punitifs qui ont globalement le même score. Toutes les distributions des scores par stratégie sont relativement moins larges car il y a moins **l'effet de hasard** lié à la composition des adversaires auxquelles un prisonnier fait face. En effet, en augmentant le nombre de journées, on **uniformise la répartition des adversaires**.

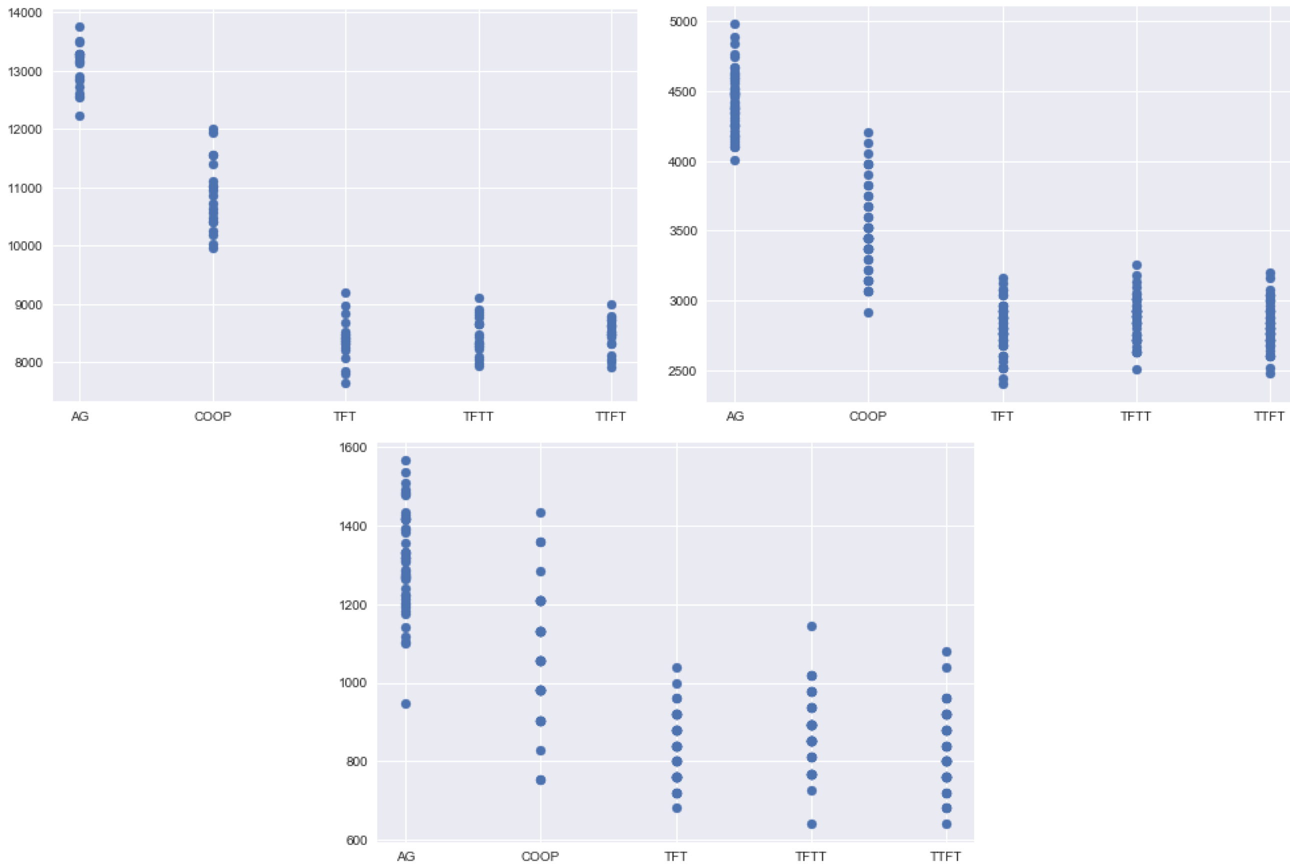

Figure 10. Comparaison avec la dimuinution du nombre de jours de duels

Les graphes ci-dessus montre l'évolution des scores avec la diminution du nombre de jours . Le graphe en haut à gauche correspond encore une fois au cas standard , le graphe en haut à droite correspond au cas 100 jours et celui du bas correspond au cas de 30 jours.

Symétriquement à ce que nous avons expliqué juste avant, on observe là aussi que globalement **les classements restent les mêmes**, mais que la **distribution des scores par stratégie s'élargit fortement** avec la baisse du nombre de jours. En effet, avec un faible nombre de journées, tous les prisonniers ne se font pas des duels entre eux, et **le hasard des adversaire** joue un grand rôle pour chacun des prisonniers. Dès lors, les résultats ne sont pas très représentatifs de l'efficacité réelle de chaque stratégie, puisque les prisonniers de chaque stratégie font chacun face à des distributions d'adversaires assez différentes.

3.5 Observation de l'impact du nombre de duels par jour

Dans le cas standard, nous avons choisi de réaliser le tournoi avec 20 confrontations par jours. Nous allons observer l'effet du nombre de confrontations par jour sur le graphe des scores par stratégie.

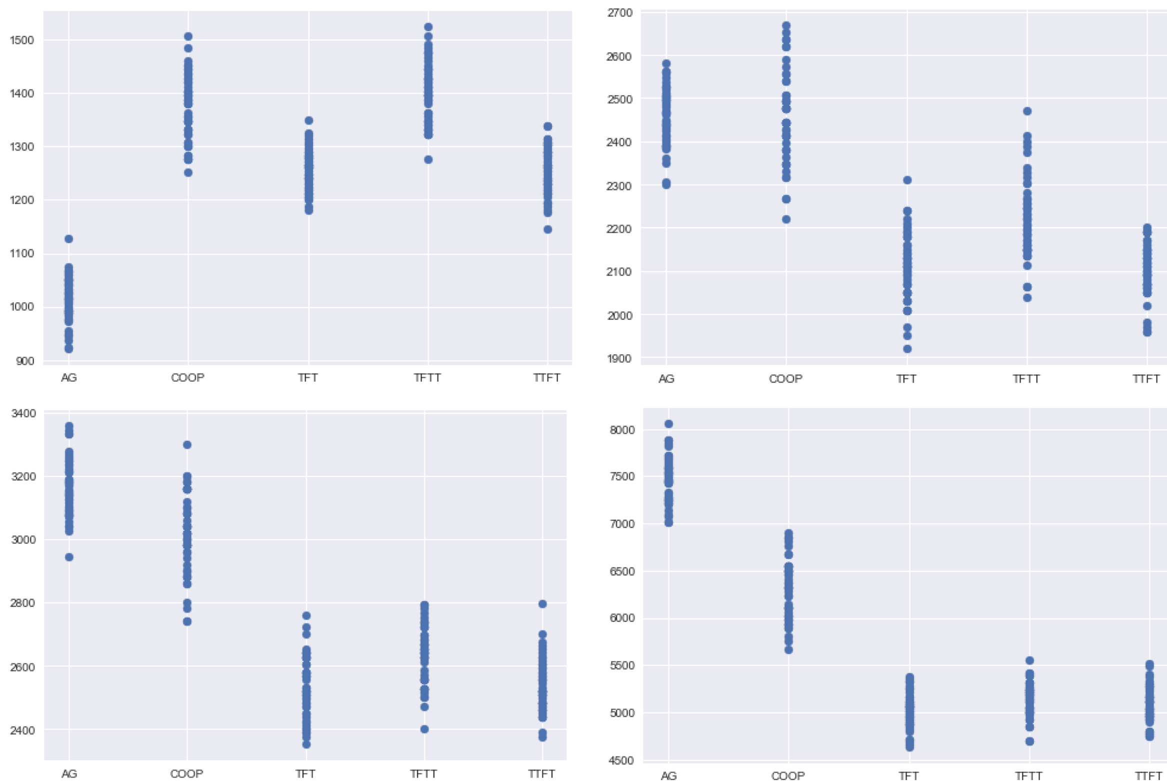

Figure 11. Graphes des scores par nombre de confrontations quotidiennes croissant

Les graphes ci-dessus montre l'augmentation du nombre de confrontations par jour. Le graphe en haut à gauche correspond au cas de 3 confrontations par jour, le graphe en haut à droite correspond au cas 5 par jour, celui en bas à gauche de 6 par jour et celui en bas à droite à 12 par jour.

On remarque que pour des valeurs inférieurs à 6, le graphe présente des distributions très différentes de ce que nous avons pu rencontrer. Ceci s'explique par le fait que ce sont **les premiers tours** des confrontations à chaque duel qui jouent **un rôle majeur** sur les scores finaux de chaque prisonnier.

Pour des valeurs inférieurs ou égales à 3 confrontations par jour, les Coopératifs et TFTT ont le même comportement et prennent de grosses condamnations lors de leurs confrontations avec les Agressifs. De même les punitifs n'ont pas encore le temps de réagir suffisemment face aux Agressifs et ont aussi des scores relativement grands.

A partir de **5 confrontations par jour**, on commence à peu près à retrouver l'ordre du cas standard à l'exception des Coopératifs qui ont encore globalement la stratégie la moins efficace à cause du fait que les Agressifs profitent encore un peu de leur avantage sur les premiers tours.

On remarque qu'il y a un **seuil vers 6 confrontations par jour**, où l'on commence à retrouver le classement des stratégies du cas standard mais avec encore des Coopératifs à la traine avec des scores aussi élevés que les Agressifs. Et plus on augmente le nombre de confrontations par jour, moins **l'effet des premiers tours** est important, et plus on observe les résultats issues du cas standard comme on peut le voir sur le dernier graphes avec 12 confrontations par jour.

Nous avons ensuite décidé d'analyser l'effet d'un très grand nombre de confrontations et voir s'il y a des changements qui s'opèrent .

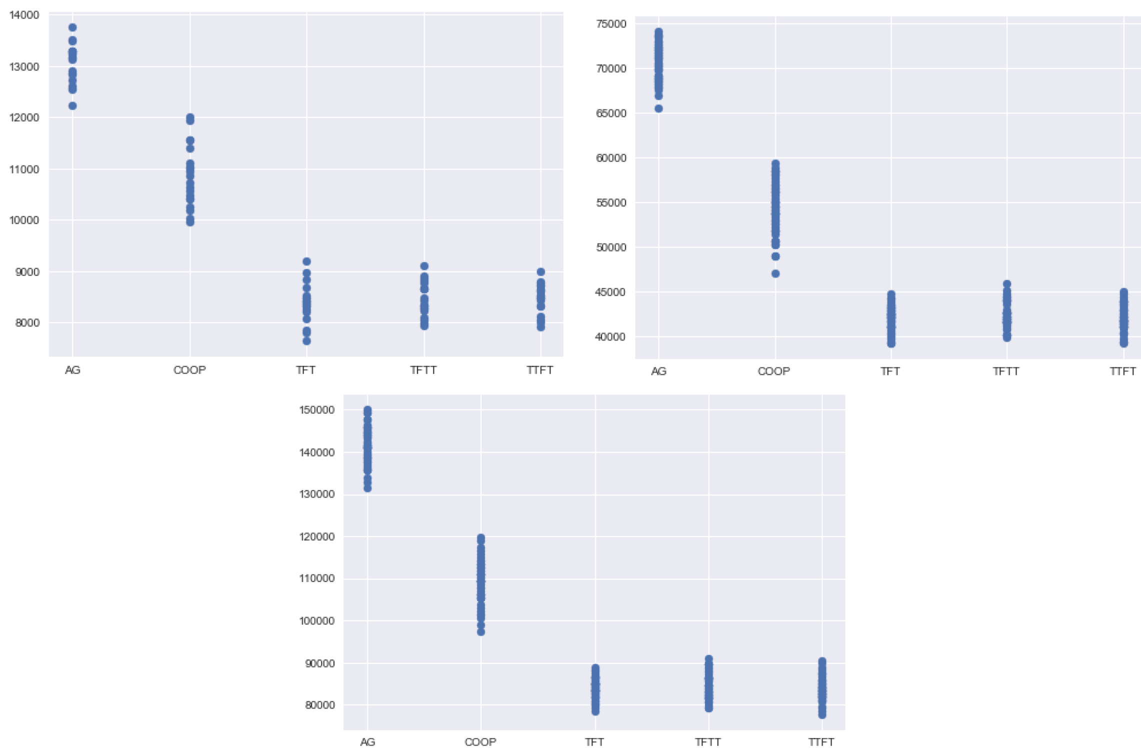

Figure 12. Comparaison avec des grandes valeurs de confrontations par jours

Les graphes ci-dessus montrent l'effet de l'augmentation du nombre de confrontrations par jour. Le graphe en haut à gauche correspond au cas standard, le graphe en haut à droite correspond au cas 100 confrontations par jour, celui du bas à 200 par jour.

On remarque que lorsqu'on augmente le nombre de confrontations par jour par rapport au cas standard, il n'y a **pas de changements significatifs** au niveau des graphes. A partir du moment qu'on dépasse un certain nombre de confrontations par jour, l'effet des premiers tours de chaque duel s'atténue, et on obtient des résultats pratiquement identiques.

La seule différence qu'on observe, c'est que lorsqu'on augmente fortement le nombre de confrontations par jours, **les distributions** des scores des prisonniers pour chaque stratégie s'aplatissent autour de la valeur moyenne.

4 EVOLUTION DES POPULATIONS

On s'intéresse pour le reste de ce rapport à l'évolution des stratégies en fonction du temps. C'est l'approche algorithme génétique qu'on adoptera pour la suite.

4.1 Architecture de l'algorithme génétique

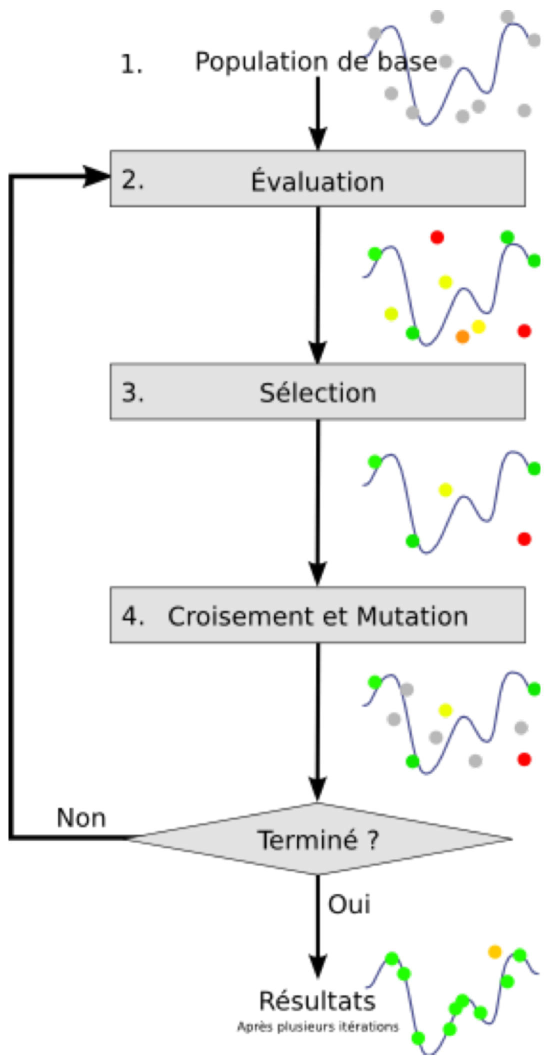

Figure 13. Architecture standard d'un algorithme génétique. D'après Wikipedia.

1. L'algorithme dépend d'une population de base
2. À chaque génération, il faut évaluer cette population
3. Il faut ensuite selectionner des " meilleurs" individus selon un critère de selection.
4. Il faut définir une fonction de reproduction et une fonction de mutation pour faire évoluer la population
5. On répète cette procédure pendant G générations.

4.1.1 • Codage

Nous avons considéré essentiellement deux approches. On peut considérer le continuum de stratégies possibles c'est-à-dire l'ensemble des matrices 2x2 à coefficients dans [0,1]. Ce codage est probabiliste car le comportement d'une stratégie dépend non seulement des deux derniers coups de l'adversaire, mais aussi du tirage d'une variable aléatoire. On peut aussi considérer un codage binaire, c'est-à-dire en prenant seulement des coefficients dans 0,1. Ce codage est déterministe car il ne dépend que des deux derniers coups de l'adversaire.

Nous avons testé les deux approches, mais nous avons décidé de garder l'approche binaire pour notre étude, car l'approche continue semblait faire converger la population très vite vers la stratégie Aléatoire, sans que nous puissions expliquer pourquoi. Il semblerait que la cause de ce phénomène soit le choix de la fonction de reproduction qui consistait à prendre la moyenne des matrices des deux parents.

Il y a 16 stratégies possibles (2^4).

4.1.2 • Population de base

Nous gardons les mêmes conventions que précédemment (J=300, D=20) Nous avons fait le choix de commencer avec une population de base similaire à celle du tournoi d'Axelrod, en prenant seulement les stratégies classiques. On choisit une population totale de 300 individus, avec 50 prisonniers de chaque stratégie et 100 prisonniers agressifs (pour compenser un peu le caractère bienveillant du reste de la population). Dans le cas contraire, on remarque leur disparition dès les deux premières itérations. On choisit de travailler sur un nombre de générations G égal à 50.

4.1.3 • Évaluation et Sélection

Pour chaque itération, on lance un tournoi sur l'ensemble de la population **(Sélection par tournoi)**, et selon le mode de reproduction, on garde une proportion p des prisonniers ayant obtenus le meilleur score (le moins d'années de prison).

4.1.4 • Croisement et mutation

Une fois la sélection effectuée, il reste à mimer une reproduction afin de générer le complément (1-p) de population manquante, afin de garder un nombre constant de prisonniers par population. À l'issu de la reproduction, il y a une probabilité de 0.05 (arbitraire) d'avoir un enfant muté. La mutation consiste à **aléatoirement changer l'un des coefficients de la matrice**.

Listing 3. Mutation dans la classe Prisonnier

```
def mutation(self):
        tirage1,tirage2=randint(0,1),randint(0,1)
        strat=self.strategy
        mat=strat.prob_mat.copy()
        mat[tirage1,tirage2]=1-self.strategy.prob_mat[tirage1,tirage2]
        new_strat=strategy(strat.coup1,strat.coup2,mat)
        self.strategy=new_strat
```

Reproduction asexuée : p=0.5

On sélectionne **la meilleure moitié** de la population et on la clone.

Listing 4. Reproduction asexuée dans la classe prisonnier

```
def asexual_reproduction(self):
        p_enfant=prisoner(self.strategy,self.generation+1)
        return(p_enfant)
```

Reproduction sexuée : p=1/3

On élimine le pire tiers de la population, et on fait reproduire les deux autres tiers. La reproduction sexuée est très similaire à la reproduction sexuée biologique : **l'enfant possède la moitié des gènes du premier parent, la moitié des gènes du deuxième** ; la répartition se fait de manière aléatoire.

Listing 5. Reproduction sexuée dans la classe prisonnier

```
def reproduction(self,p2):
        tirage1,tirage2=randint(0,1),randint(0,1)
        tirage3,tirage4=randint(0,1),randint(0,1)
        while (tirage3,tirage4)==(tirage1,tirage2):
            tirage3,tirage4=randint(0,1),randint(0,1)
        mat_enfant=self.strategy.prob_mat
        mat_enfant[tirage1,tirage2]=p2.strategy.prob_mat[tirage1,tirage2]
        mat_enfant[tirage3,tirage4]=p2.strategy.prob_mat[tirage3,tirage4]
        strat=strategy(self.strategy.coup1,p2.strategy.coup2,mat_enfant)
        p_enfant=prisoner(strat,self.generation+1)
        return(p_enfant)
```

4.1.5 • Représentation des résultats

Au fil des générations, des stratégies nouvelles apparaissent par croisement/mutation, et vient l'enjeu de pouvoir **représenter l'évolution de ces populations au cours du temps**. Nous avons décidé de comptabiliser d'une part **le nombre total de trahisons par génération** comme une mesure de l'agressivité globale d'une population.

Ensuite, nous nous sommes intéressés au coefficient (0,0) d'une stratégie, qui représente la probabilité de trahison sachant que le complice a coopéré aux deux derniers rounds. Si ce coefficient est égal à 0, le prisonnier est considéré comme « bienveillant ». On comptabilise **le nombre total de prisonniers bienveillants pour chaque génération**, ce qui nous donne une mesure de la bienveillance globale de la population. On s'attend à ce que cet indicateur soit **négativement corrélé au premier**.

Enfin, nous avons naturellement tenté de représenter l'évolution temporelle de la fréquence de chaque stratégie possible, cependant aucune tendance ne semblait en ressortir au fil des simulations. Les stratégies apparaissent et disparaissent très vite, et avoir 16 courbes de 16 couleurs différentes sur un graphique qui fluctuent en permanence apporte plus de confusion que de clarification.

Nous nous contenterons d'afficher la répartition finale des stratégies, ainsi que les deux premiers indicateurs en fonction du temps.

4.2 Différents scenarios de dynamique

Au fil des simulations, nous avons remarqué qu'il n'y avait **pas en général de convergence** vers une stratégie en particulier ou même un équilibre. Il y a en effet plusieurs scenarios possibles, régis par le hasard dû aux mutations et à la reproduction, ainsi qu'à l'apparition de certaines nouvelles stratégies pour le moins **originales**.

4.2.1 • Reproduction asexuée

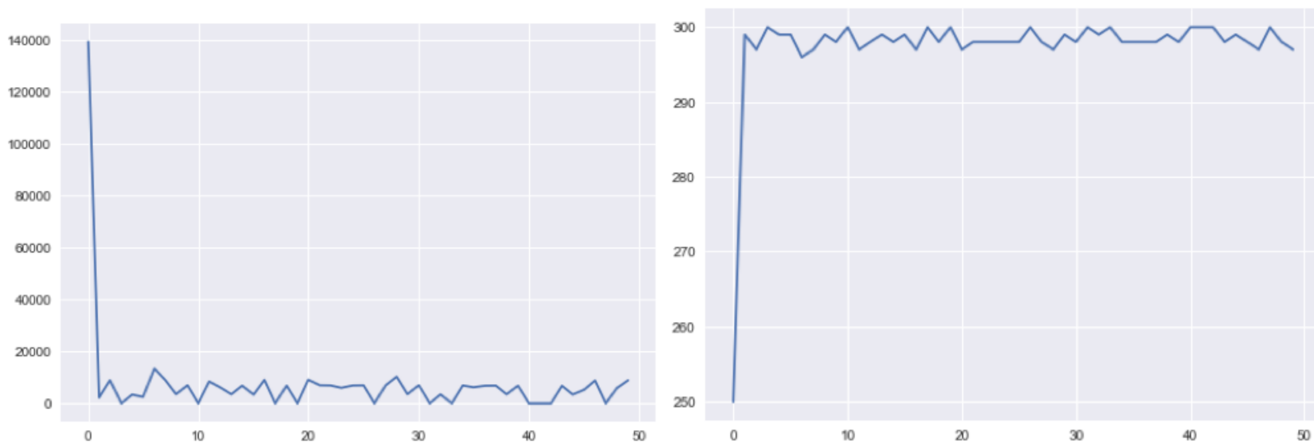

Figure 14. Agressivité(à gauche) et bienveillance(à droite) en fonction du temps pour un mode de reproduction asexuée

Pour le mode de reproduction asexué, il semble clair que l'aléatoire joue très peu. Seules les mutations apportent un plus d'aléa. En quelques générations, toutes les stratégies agressives se font éliminer (d'autant plus que la sélection est stricte, seuls les 50% meilleurs survivent), les stratégies punitives mais bienveillantes prennent de plus en plus de terrain (par clonage) et on arrive rapidement à un nombre de trahisons très faible. Le nombre de stratégies bienveillantes (qui ne trahissent pas les complices bienveillants deux fois de suite) converge vers son maximum (300) dès les premières générations. On a pour cette dernière simulation la répartition finale suivante des stratégies :

La stratégie dominante est TitForTat, suivie par TwoTitsForTat. On a l'apparition de nouvelles stratégies, par exemple :

$$\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$

Figure 15. Nouvelle stratégie

Un prisonnier doté de cette stratégie trahit seulement si l'adversaire a trahi au tour n-2, oublie totalement le tour précédent. Cette stratégie très originale est cependant peu représentée.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Figure 16. Nouvelle stratégie

Un prisonnier doté de cette stratégie trahit si l'adversaire a trahi exactement une fois au cours des deux derniers tours, mais pardonne s'il trahit deux fois de suite. Cette stratégie, très surprenante, qui est globalement bienveillante réussit à survivre car le reste de la population est globalement bienveillant.

Remarquons que les stratégies totalement agressives et totalement coopératives ont disparu de la population.

Pour conclure, dans le cas d'une reproduction asexuée, on a une **convergence très rapide vers un équilibre optimal de bienveillance**, avec très peu de trahisons et une coopération globale. Cependant, ce modèle n'évolue en fait très peu. Il ne fait que sélectionner les meilleures stratégies parmi celles que l'on connaît déjà. En d'autres termes, il ne laisse pas lieu à la diversité de la population, et il y a très peu de surprises.

4.2.2 • Reproduction sexuée : Des scenarios plus divers

Passons maintenant au modèle de reproduction sexuée. Au fil des simulations, nous avons remarqué que le modèle ne convergeait pas systématiquement vers la solution optimale globale (bienveillance généralisée). La raison de cette diversité est le caractère aléatoire de la reproduction sexuée. Il y a en fait plusieurs scénarios possibles d'évolution, qui se répartissent essentiellement en trois catégories.

Scenario 1 : Un équilibre optimal relativement stable

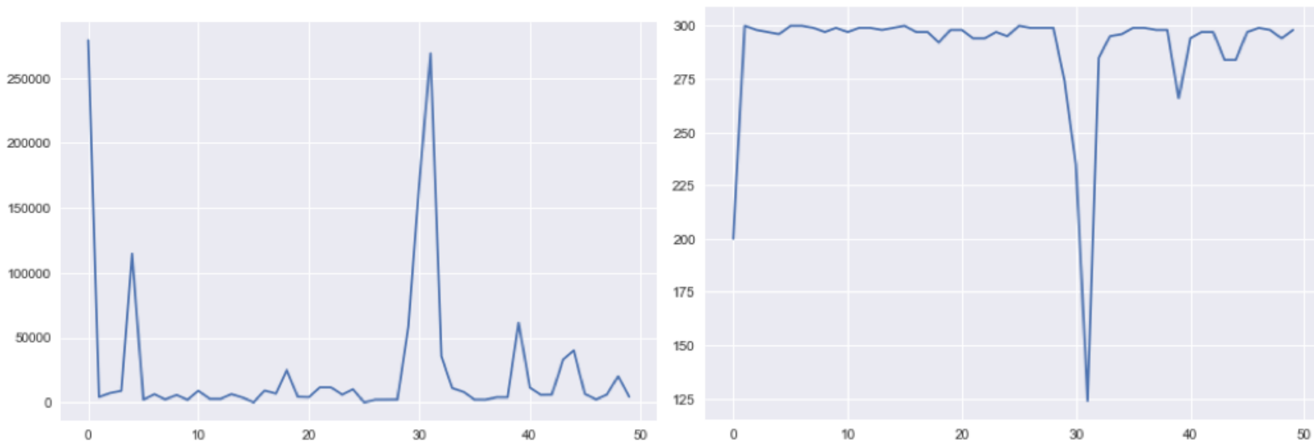

Figure 17. Agressivité(à gauche) et bienveillance(à droite)en fonction du temps pour un mode de reproduction sexuée.

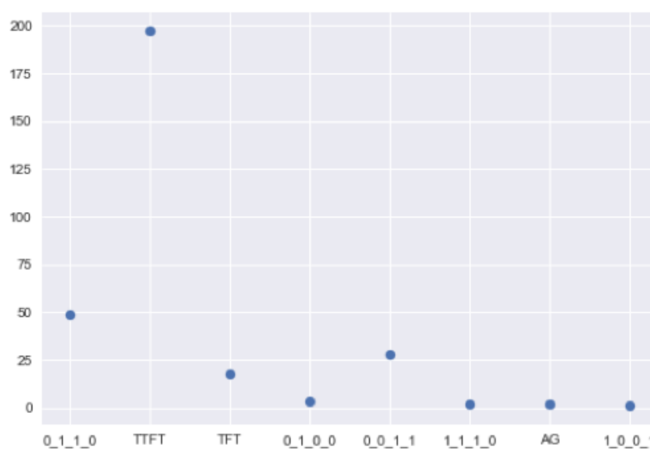

Figure 18. Répartition des stratégies survivantes

Ce scenario décrit une dynamique d'évolution qui **converge vers la solution optimale globale**, mais avec quelques perturbations (que nous appelons « **Révolutions agressives** ») .

Au départ, la sélection fait diminuer l'agressivité globale de la population. Si la proportion d'individus agressifs ne dépasse pas un certain seuil, ils sont systématiquement éliminés à la fin du tournoi. Ce n'est que lorsque cette proportion excède ce seuil qu'il y a une Révolution agressive, pendant laquelle les individus plus bienveillants ne sont pas assez nombreux pour collaborer et gagner des points entre eux, et se font écraser par les individus plus agressifs.

Dans ce scenario, cette situation ne dure pas, car l'aléatoire lié à la reproduction rééquilibre la population en créant des individus punitifs, qui collaborent avec les individus bienveillants naïfs et punissent les individus trop agressifs.

Scenario 2 : La victoire de l'agression

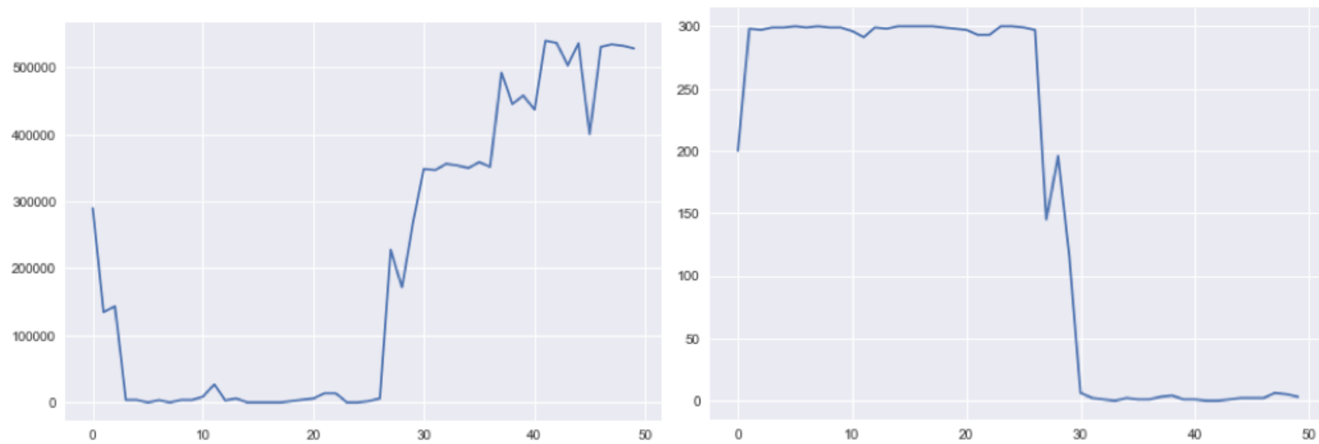

Figure 19. Agressivité(à gauche) et bienveillance(à droite)en fonction du temps pour un mode de reproduction sexuée.

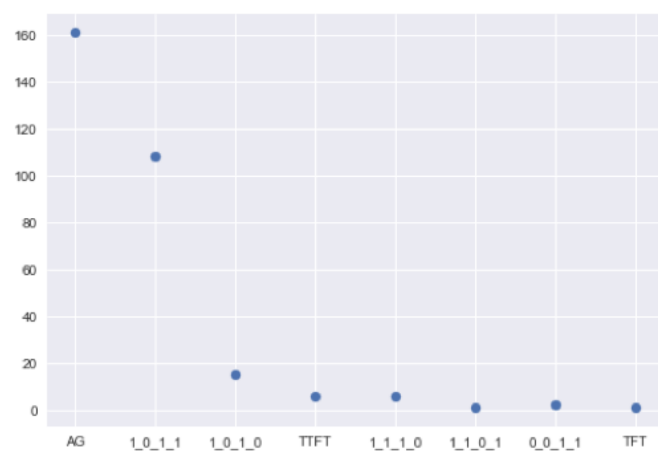

Figure 20. Répartition des stratégies survivantes

Ce scenario décrit une dynamique d'évolution qui converge vers une population globalement agressive, avec un nombre de trahisons par génération proche de son maximum (Les prisonniers se dénoncent mutuellement). C'est un exemple de Révolution agressive réussie. Au départ, la sélection fait baisser drastiquement le nombre de trahisons, mais l'aléatoire lié à la reproduction et aux mutations fait apparaître un grand nombre d'individus plus agressifs. La proportion de la population qui est punitive et bienveillante ne suffit pas à contrebalancer la Révolution : Les agressifs gagnent la bataille et représentent la quasi-majorité de la population. Il est alors très difficile de les renverser. En effet, les individus punitifs/bienveillants n'ont de bons scores que lorsque la proportion d'individus bienveillants est assez grande pour leur permettre d'amasser des points. Ceux, en revanche, qui sont naïvement bienveillants, se font complètement écraser et ne survivent pas.

Les stratégies survivantes dominantes sont : la stratégie Agressive en tête, suivie de la nouvelle stratégie :

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Figure 21. Nouvelle stratégie

Un prisonnier doté de cette stratégie trahit systématiquement sauf lorsque l'adversaire a trahi au dernier et pas au tour d'avant. Cette stratégie surprenante, globalement agressive, survit et permet surtout de perpétuer les gènes agressifs.

Scenario 3 : Un équilibre instable

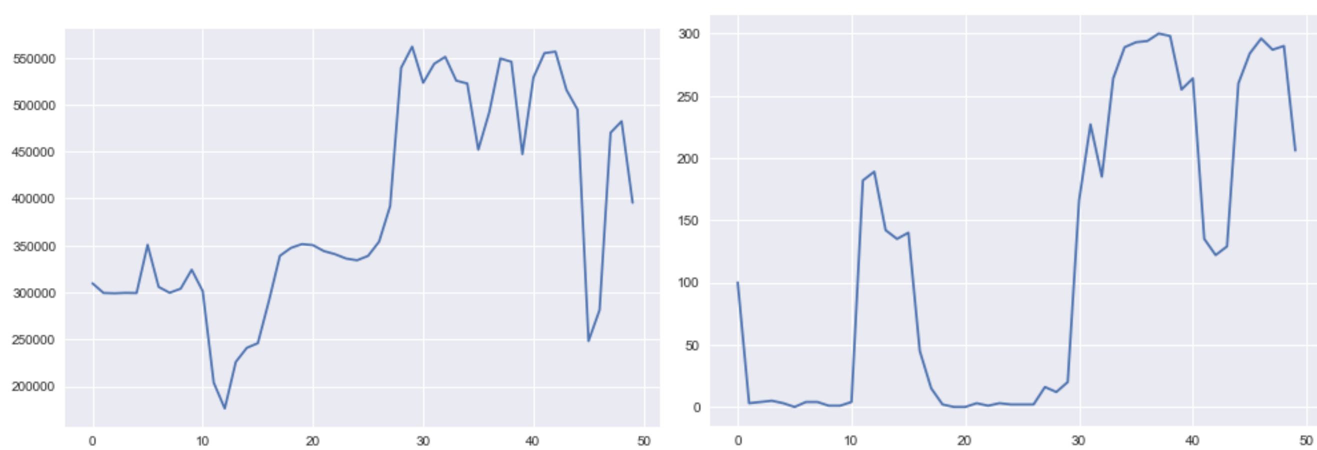

Figure 22. Agressivité(à gauche) et bienveillance(à droite)en fonction du temps pour un mode de reproduction sexuée.

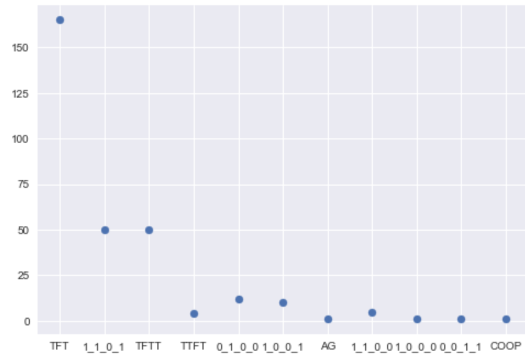

Figure 23. Répartition des stratégies survivantes

Ce scenario décrit une dynamique d'évolution qui ne semble pas converger. Elle oscille en cycles, entre des phases pendant lesquelles les stratégies bienveillantes dominent et d'autres phases pendant lesquelles les stratégies agressives dominent.

Notons que ce scenario est un peu plus fréquent que les deux autres. Notons aussi que lorsque nous augmentons l'échelle de temps (autrement dit lorsqu'on augmente le nombre de générations, cf figure 23), la dynamique d'évolution semble presque systématiquement être celle-ci, et **les cas de convergence sont très rares**.

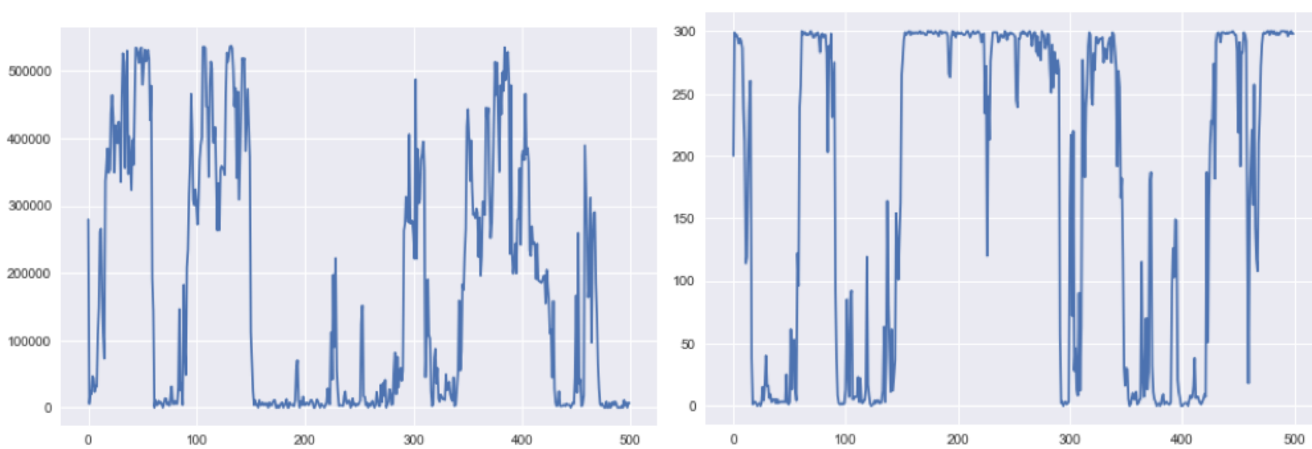

Figure 24. Agressivité(à gauche) et bienveillance(à droite)en fonction du temps pour G=500.

5 CONCLUSION

En conclusion, ce projet de recherche nous a permis d'appréhender un des exemples majeurs de la théorie des jeux. Le dilemme du prisonnier est la modélisation d'une situation qui est fréquemment rencontré dans le monde des sciences et des interactions humaines. Nous avons pu voir qu'en théorie, la solution d'équilibre en stratégie dominante qui est aussi la solution de l'équilibre de Nash est d'agir d'une manière agressive dans le cas où le jeu se déroule en un nombre N de tours connus par les joueurs. Dans le cas où le nombre de tours est une variable inconnue pour les joueurs, nous avons simulé le tournoi d'Axelrode classique avec 5 stratégies de base et analysé lesquelles étaient les plus efficaces en faisant varier un certain nombre de paramètres. On retient que globalement **les stratégies les plus efficaces sont les stratégies punitives** qui s'adaptent au comportement de l'adversaire : elles profitent des avantages de la coopération lorsque l'adversaire est coopératif et contre attaquent contre les adversaires agressifs et optimisent ainsi le résultat de leurs duels.

Dans un second temps, nous avons utilisé le tournoi d'Axelrod pour selectionner les meilleurs individus d'une génération, dans l'espoir d'aboutir à une solution optimale au dilemme du prisonnier. Nos simulations ne nous permettant pas de conclure à une solution optimale avec notre modélisation, elles suggèrent néanmoins un caractère cyclique, alternant coopération globale et agressivité globale (équilibre global optimal et équilibre de Nash), sans jamais converger.

Il n'y a pas pour le moment de résultat théorique qui offre une solution unique pour des configurations du jeu mais nous avons pu observer divers comportements et dynamiques qui peuvent être interpretés de plusieurs manières. D'autres pistes ou facteurs auraient pu être exploités lors de ce projet de recherche. Nous aurions pu essayer par exemple de trouver pour chacune des stratégies un adversaire optimal et un pire ennemi pour mieux comprendre les différents scenarios. Nous aurions pu tenter de simuler les tournois en faisant varier les matrices de gains mais lors de nos expériences, nous avons remarqué que les résultats étaient sensiblement identiques et que cela aurait nécessité beaucoup plus de temps de simulations sans résultat nettement différent. Enfin, nous aurions pu approfondir notre recherche de stratégies non déterministes avec des coefficients non binaire pour observer les résultats des matrices de réactions continues.

REFERENCES

[1] Robert Axelrod and William D. Hamilton : The Evolution of Cooperation. Science, New Series, Vol. 211, No. 4489. (Mar. 27, 1981), pp. 1390-1396.

[2] https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/game-theory/axelrod.html

[3] http://www.aunege.org/ressources/pdf/dilemmeprisonnier.pdf