Intelligence artificielle & intelligence collective

Théorie des jeux

Laboratoire d'Informatique Fondamentale de Lille

Année 2003-2004

Bruno Beaufils
(beaufils@lifl.fr)

Partie A

Introduction à la théorie des jeux

LIFL

Historique

- Borel (1921), Von Neumann (1928)
- «*Theory of Games and Economic Behaviour*», Von Neumann et Morgenstern (1942)
- Exemples de *jeux* :
  - Jeux de société (échecs, dames, go, etc), Jeux de cartes (bridge, poker, etc)
  - Enchères
  - Partages de ressource (marchandage)
  - Fixation du prix d'un bien dans un marché compétitif
  - Est-ce que j'écoute de la musique ce soir ?
  - ...

Introduction

- Jeu : situation où des individus sont conduits à faire des choix parmi un certain nombre d'actions possibles, et dans un cadre défini à l'avance.
  ➠ Situation de conflit.
- Objectif : étude des comportements rationnels des individus en situation de conflit.
- Ses résultats peuvent avoir des implications dans des domaines importants : économie, biologie, science sociale, science politique, etc.
- Trois approches en théorie des jeux :
  - Théorie des jeux classique.
  - Théorie des jeux évolutionnaire.
  - Théorie des jeux computationnelle.

Définition

La théorie des jeux s'intéresse aux problèmes posés par l'interaction stratégique d'agents rationnels poursuivant des buts qui leur sont propres.

- ☞ Étude d'interactions
- ☞ Manipulation d'agents *rationnels* (maximisant une fonction)
- ☞ 2 approches :
  1. Descriptive
  2. Normative

Approche descriptive

– Eric Rasmussen :

«*C'est là exactement le paradigme de la théorie des jeux : celui qui construit le modèle attribue des fonctions de gain et des stratégies aux joueurs, puis* observe *ce qui se passe lorsqu'ils choisissent des stratégies pour obtenir le gain maximum*».

– Ken Binmore :

«*La théorie des jeux, telle qu'elle est développée actuellement, est surtout la* description *de ce qui se passe lorsque des personnes interagissent rationnellement*».

– David Kreps :

«*l'objet de la théorie des jeux est d'aider les économistes à* comprendre *et à* prédire *ce qui se produit dans différentes situations économiques*».

Approche normative

– Eric Van Damme :

> «*Game Theory is a normative theory : it aims to prescribe what each player in a game should do in order to promote his interests optimally*».

– Robert Sugden :

> «*My approach, like that of classical game theory, will be normative : I shall try to show why and how it might be rational for players to make use of the information provided by labels*».

– Luce et Raiffa :

> «*Il est essentiel, pour nous, que le chercheur en sciences humaines sache que la théorie des jeux n'est pas* descriptive, *mais plutôt (conditionnellement)* normative. *Elle n'établit ni comment les gens se comportent, ni comment ils devraient le faire pour atteindre certains buts. Elle prescrit, avec des hypothèses données, des types d'action qui conduisent à des issues ayant un certain nombre de propriétés qui relèvent de l'optimalité*».

Définition à retenir

– Encyclopedia Britannica :

«*A solution to a game* prescribes *the decision the players should make and* describes *the game's appropriate outcome. Game theory serves as a guide for players and as a tool for* predicting *the outcome of a game*».

– Bernard Guerien :

«*Selon l'acceptation courante, un jeu est une situation où des individus (les* joueurs*) sont conduits à faire des choix parmi un certain nombre d'actions possibles, et dans un cadre défini à l'avance (les* règles du jeu*), le résultat de ces choix constituant une issue du jeu, à laquelle est associé un gain, positif ou négatif, pour chacun des participants.*»

Utilité (1)

☞ Une hypothèse de base de la théorie des jeux est de considérer que les agents sont rationnels, c'est-à-dire qu'ils tentent d'arriver à la situation la *meilleure* pour eux.

☞ On appelle *utilité* la mesure de chaque situation aux yeux de l'agent.

☞ L'*utilité* n'est pas une mesure du gain matériel, monétaire, etc. mais une mesure *s*ubjective du contentement de l'agent.

La fonction d'utilité lie un ordre de préférences à des valeurs numériques.

➠ Les valeurs utilisées par la fonction n'ont pas d'importance, seul l'ordre des préférences en a.

Utilité (2)

Supposons qu'un agent x préfère une situation a à une situation b et une situation b à une situation c.

Une fonction d'utilité valide peut être :

- ☞ $a \mapsto 3$
- ☞ $b \mapsto 2$
- ☞ $c \mapsto 1$

Une autre fonction **équivalente** au sens de la théorie des jeux :

- ☞ $a \mapsto 10\ 354$
- ☞ $b \mapsto 0$
- ☞ $c \mapsto -10$

On ne mesure pas une quantité mais un ordre.

Représentation d'un jeu

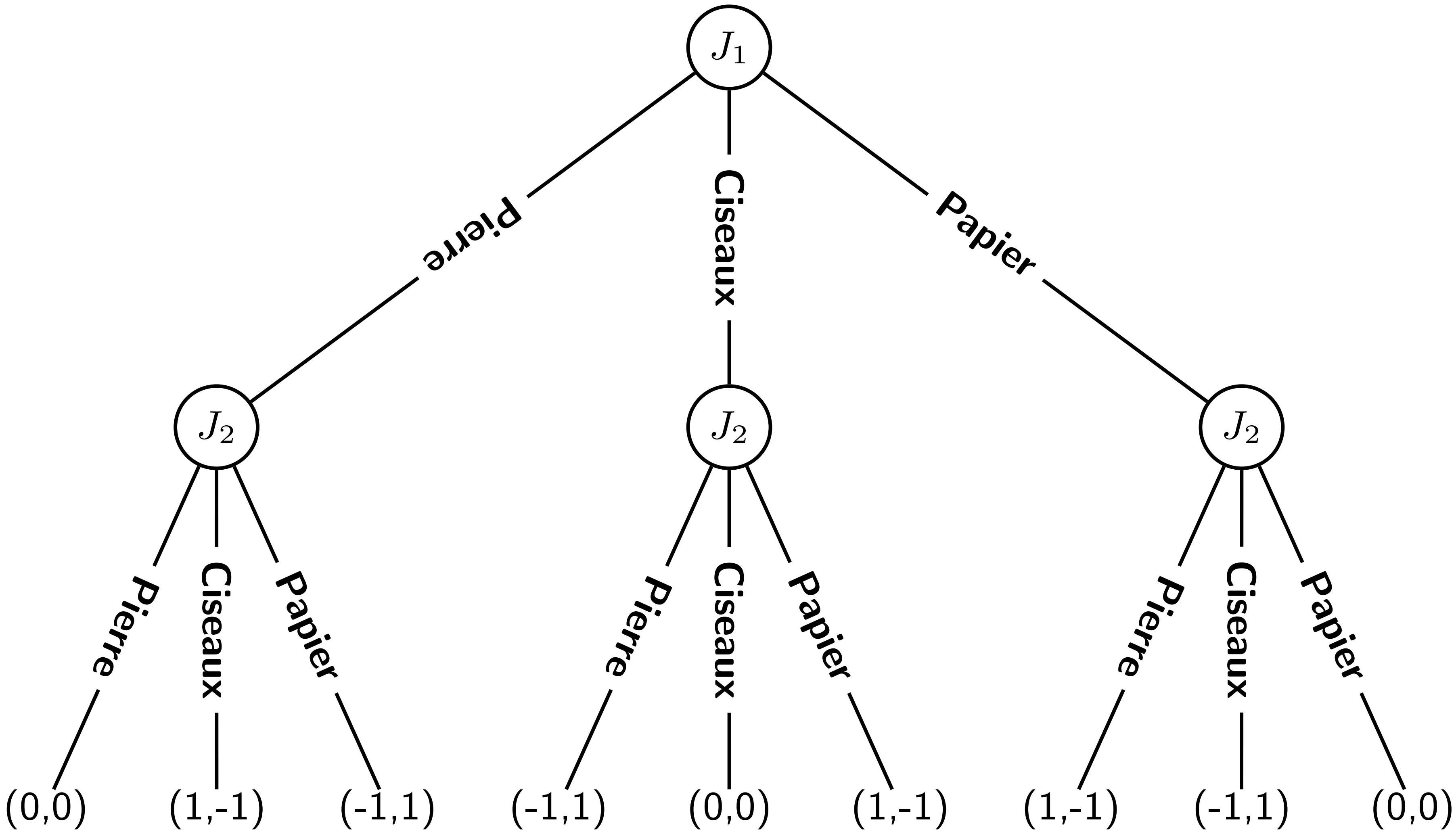
Il existe 2 formes de jeu :

① Forme extensive
Utilisation d'un arbre

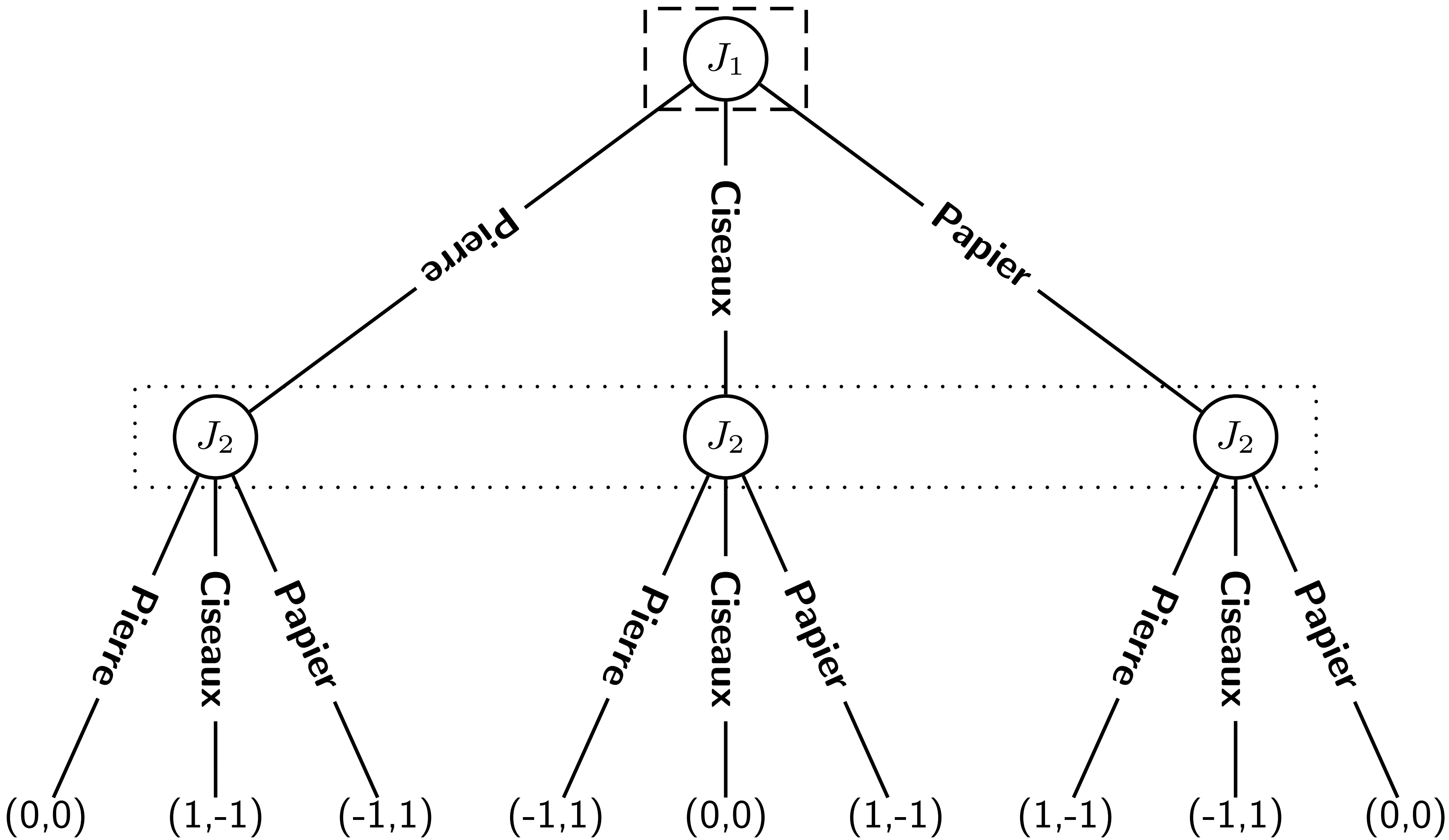
② Forme stratégique (forme normale)
Utilisation d'une matrice

- A chaque jeu sous forme extensive correspond un jeu sous forme stratégique dans lequel les joueurs choisissent simultanément les stratégies qu'ils mettront en oeuvre.
- En revanche, un jeu sous forme stratégique peut correspondre à plusieurs jeux sous forme extensive différents.

Forme extensive (1)



Forme extensive (1)

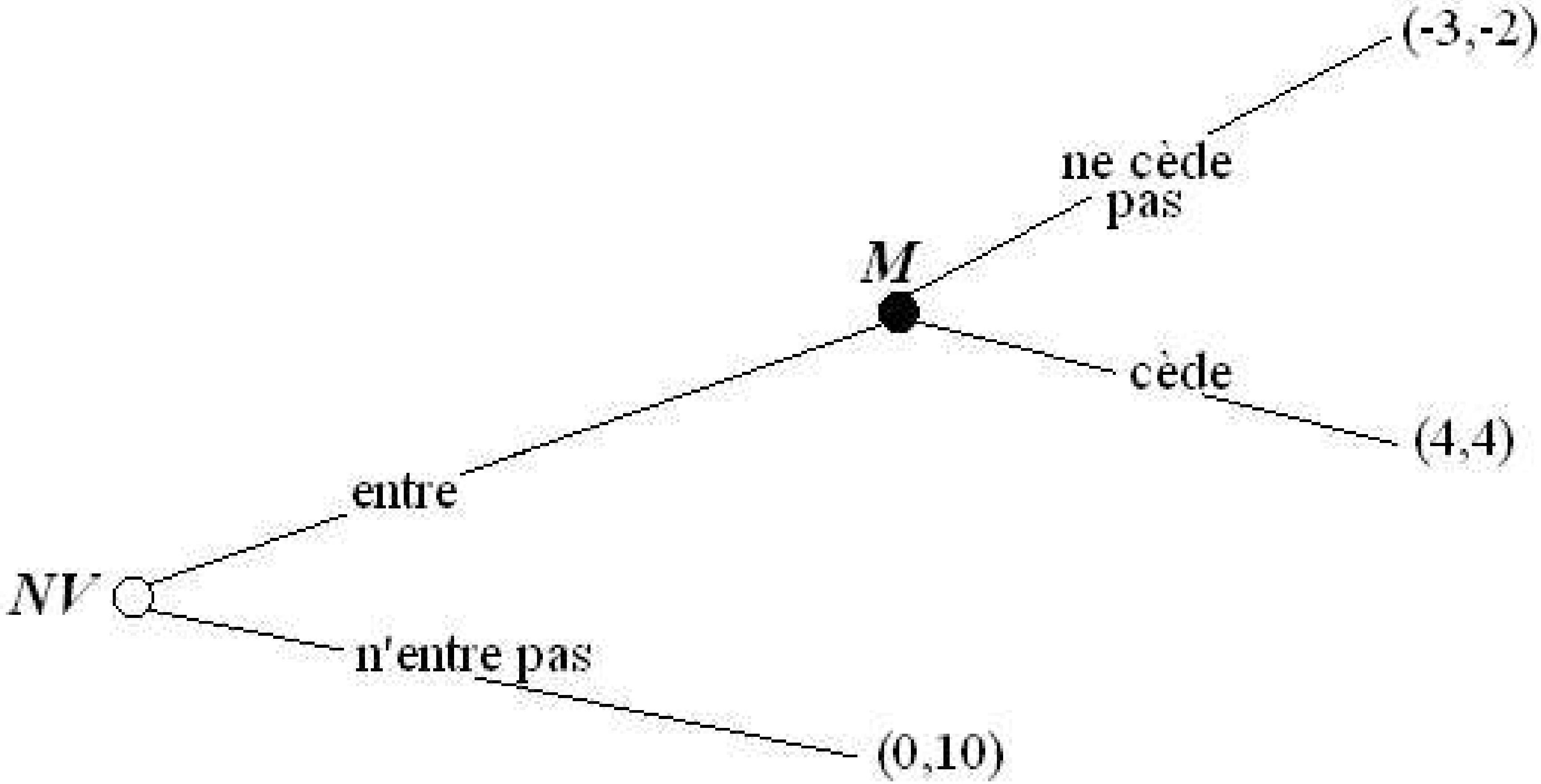


Forme stratégique (1)

J_2

J_1		**Pierre**	**Ciseaux**	**Papier**
	Pierre	(0,0)	(1,-1)	(-1,1)
J_1	**Ciseaux**	(-1,1)	(0,0)	(1,-1)
	Papier	(1,-1)	(-1,1)	(0,0)

Forme extensive (2)



Forme stratégique (2)

	M	
NV	**cède**	**ne cède pas**
entre	$(4,4)$	$(-3,-2)$
n'entre pas	$(0,10)$	$(0,10)$

Stratégie

☞ Dans une forme extensive, une *stratégie* est la spécification complète du comportement d'un joueur dans n'importe quelle situation

☞ Dans une forme stratégique, une *stratégie* correspond au choix d'une ligne ou d'une colonne

➠ Une stratégie correspond à un comportement

Récurrence à rebours

Élimination des stratégies dominées

Joueur 2

Joueur 1

	u	v
x	4,2	3,1
y	2,5	9,0

Élimination des stratégies dominées

Joueur 2

Joueur 1

	u	v
x	4,2	3,1
y	2,5	9,0

Élimination des stratégies dominées

Joueur 2

Joueur 1

	u	v
x	4,2	3,1
y	2,5	9,0

Équilibre de Nash

Joueur j

		G	A
Joueur i	G	(3,3)	(0,5)
	A	(5,0)	(1,1)

Équilibre de Nash

		Joueur j	
		G	A
Joueur i	G	(3,3)	(0,5)
	A	(5,0) $\times$	(1,1)

Équilibre de Nash

		Joueur j	
		G	A
Joueur i	G	(3,3)	(0,5)
	A	(5,0) $\times$	(1,1) $\times$

Équilibre de Nash

		Joueur j	
		G	A
Joueur i	G	(3,3)	(0,5) $\circ$
	A	(5,0) $\times$	(1,1) $\times$

Équilibre de Nash

		Joueur j	
		G	A
Joueur i	G	(3,3)	(0,5) $\circ$
	A	(5,0) $\times$	(1,1) $\times$ $\circ$

Équilibre de Nash

		Joueur j	
		G	A
Joueur i	G	(3,3)	(0,5) $\circ$
	A	(5,0) $\times$	(1,1) $\times$ $\circ$

➥ L'équilibre de Nash est (A, A).

Stratégies pures/Stratégies mixtes

☞ Les stratégies que nous avons définies et utilisées pour le moment sont des *stratégies pures*, c'est-à-dire les options qui se présentent aux joueurs.

☞ Une *stratégie mixte* σ_i est une distribution de probabilité sur l'ensemble des stratégies pures.

☞ L'ensemble des stratégies pures utilisées (i.e. dont la probabilité n'est pas nulle) par une stratégie mixte σ_i est appelé le *support* de la stratégie mixte.

➠**Théorème de Nash** : *Tout jeu sous forme stratégique a un équilibre de Nash en stratégies mixtes.*

Une petite taxonomie

- Jeux à somme nulle (strictement compétitifs) / Jeux à somme non-nulle
- Jeux à information complète / Jeux à information incomplète
- Jeux à information parfaite / Jeux à information imparfaite
- Jeux coopératifs / Jeux non-coopératifs
- Jeux à 2 joueurs / Jeux à n joueurs

		Pol.	
		main droite	main gauche
Pri.	main droite	$(1, -1)$	$(-1, 1)$
	main gauche	$(-1, 1)$	$(1, -1)$

Partie B

Le dilemme itéré du prisonnier

LIFL

Historique

- Modèle de la théorie des jeux introduit à la RAND Corp par Flood et Dresher (1952)
  - L'objectif est de prendre en défaut la théorie des jeux
- Énoncé sous l'histoire des prisonniers par Tucker
- Popularisé au début des années 80 par Axelrod
- Très souvent **utilisé** (200 publications entre 1988 et 1994) notamment en économie, psychologie et en biologie théorique
- Étudié surtout dans des espaces continus et non déterministes
- Le modèle simple, souvent considéré comme trop théorique, a souvent été délaissé pour des modèles plus *proches* de la réalité.

L'histoire

Deux personnes arrêtées ensemble en possession d'armes à feu sont soupçonnés d'un délit fait en commun. Les policiers les séparent et disent à chacun :

- ☞ Si un des deux avoue et que l'autre n'avoue rien, le premier est libéré, et le second emprisonné (5 ans) ;
- ☞ Si les deux avouent, les deux iront en prison (3 ans) ;
- ☞ Si aucun des deux n'avoue, les deux seront seront libérés assez vite (1 an).

Vous êtes un des deux prisonniers, que faites-vous ?

Un autre cas

Vous n'avez pas vraiment les mêmes goûts que votre voisin en matière de musique. Il lui arrive souvent d'écouter de la techno à fond. De même il vous arrive (en représailles) de mettre votre musique à un volume plus que raisonnable. Ce qui a pour conséquences que le lendemain il recommence à nouveau. En dehors de ces périodes agitées, vous appréciez les périodes où aucun de vous ne gêne l'autre.
Supposons que l'on pondère votre satisfaction :

- ☞ Vous avez une satisfaction de **5** à écouter votre musique à un volume important.
- ☞ La satisfaction est de **0** lorsque votre voisin met sa musique à fond.
- ☞ Une soirée "calme", sans musique vous apporte une satisfaction de **3**.
- ☞ Le fait d'écouter "simultanément" votre musique mêlée à celle du voisin, donne une satisfaction de **1**.

Vous savez ce que votre voisin à eu comme comportement les jours précédents, que faites-vous aujourd'hui ?

Dilemme du prisonnier. . .

☞ Au sens de la théorie des jeux c'est un jeu :
- simultané
- symétrique
- à deux joueurs
- à somme non nulle
- non coopératif

☞ chaque joueur doit choisir une parmi deux cartes :
- la *Coopération* (C)
- la *Trahison* (D)

☞ Il y a dilemme car l'*intérêt individuel* (≡ la rationalité au sens économique) (D,D) rapporte moins que l'*intérêt collectif* (C,C)

L'équilibre de Nash c'est (D,D)

Matrice de gain

La matrice est symétrique, on ne montre donc que le score du joueur ligne.

	Cooperate	Defect
Cooperate	$R = 3$ *R*eward récompense pour coopération mutuelle	$S = 0$ *S*ucker's payoff salaire de la dupe
Defect	$T = 5$ *T*emptation Tentation à trahir	$P = 1$ *P*unishment punition pour la trahison mutuelle

$$S < P < R < T$$

. . .itéré

☞ les joueurs se recontrent un certain nombre de fois

☞ les joueurs ne connaissent pas le terme du jeu

☞ le gain d'un joueur est le cumul de ses gains dans chaque rencontre

☞ Pour favoriser la coopération on rajoute la contrainte :

$$S + T < 2R \tag{1}$$

Quelques situations modélisées

☞ Deux pays doivent-ils lever des taxes douanières sur les produits importés de l'autre pays.

☞ Deux entreprises concurrentes doivent-elles essayer de s'entendre pour se partagé un marché ou se faire concurrence ?

☞ Deux espèces vivant sur un même territoire doivent-elles cohabiter ou se disputer la nourriture disponible ?

Stratégies

- À chaque étape un joueur sait ce que son adversaire a joué dans les coups précédents
- Il est alors possible de définir un **comportement** prédéfini pour le jeu : une **stratégie**
- Quelques exemples de stratégies :

  - `all_c`
  - `all_d`
  - `per_ccd`
  - `soft_majo`
  - `tit_for_tat`
  - `spiteful`
  - `ipd_random`
  - `gradual`

⇒ Identification d'un agent (un joueur) à sa stratégie (son comportement)

Stratégies (exemple)

	1	2	3	4	5	6	7	8	9	10		
	3	3	0	3	3	0	3	3	0	3	=	21
`all_c`	C	C	C	C	C	C	C	C	C	C		
`per_ccd`	C	C	D	C	C	D	C	C	D	C		
	3	3	5	3	3	5	3	3	5	3	=	36

	0	1	1	1	1	1	1	1	1	1	=	9
`tit_for_tat`	C	D	D	D	D	D	D	D	D	D		
`all_d`	D	D	D	D	D	D	D	D	D	D		
	5	1	1	1	1	1	1	1	1	1	=	14

Quelle est la meilleure stratégie ?

- ☞ **Celle qui bat toutes les autres** ?
  `all_d`
- ☞ **Celle qui fasse le meilleur score possible face à toutes les autres** ?
  aucune, car faire le score maximal contre `all_d` et contre `spiteful` est impossible

➠ Problème de la définition d'un critère d'évaluation de stratégies

Sur des confrontations de 100 parties, le gain maximal est de 500 points et le gain minimal est de 0 point
➠ C'est ce qu'obtiennent `all_d` et `all_c` l'une contre l'autre.

Mais...
- ☞ 2 gentilles entre elles obtiennent chacune 300 points
- ☞ 2 méchantes entre elles obtiennent chacune 100 points

Chaque stratégie est bonne (au sens du meilleur score) face à certaines et mauvaises face à d'autres car elle ne sait pas à qui elle a affaire.

Évaluations de stratégies

Il existe deux grands types de méthodes d'évaluation :

1. **Le tournoi**
   - Des stratégies se rencontrent deux à deux comme dans un championnat de *football*
   - Le score d'une stratégie est la somme des scores de tous ses matchs
   - matchs de même longueur inconnue des joueurs
   - chaque stratégie est classée en fonction de son score : $\sum V(i|j)$
2. **Les évolutions écologiques**
   - Population polymorphe d'individus : chaque stratégie est représenté par N individus dans une population
   - A tournoi (round-robin) entre tous les individus est fait
   - Les stratégies *faibles* sont désavantagées alors que les **bonnes** sont favorisées par une redistribution proportionnelle des stratégies sur les individus
   - Le facteur d'adaptation d'un individu est la somme des scores obtenus face à tous les autres individus
   - Ce cycle est répété jusqu'à une stabilisation de la population
   - *Pas de mutation, ➠ ça n'est pas un algorithme génétique*

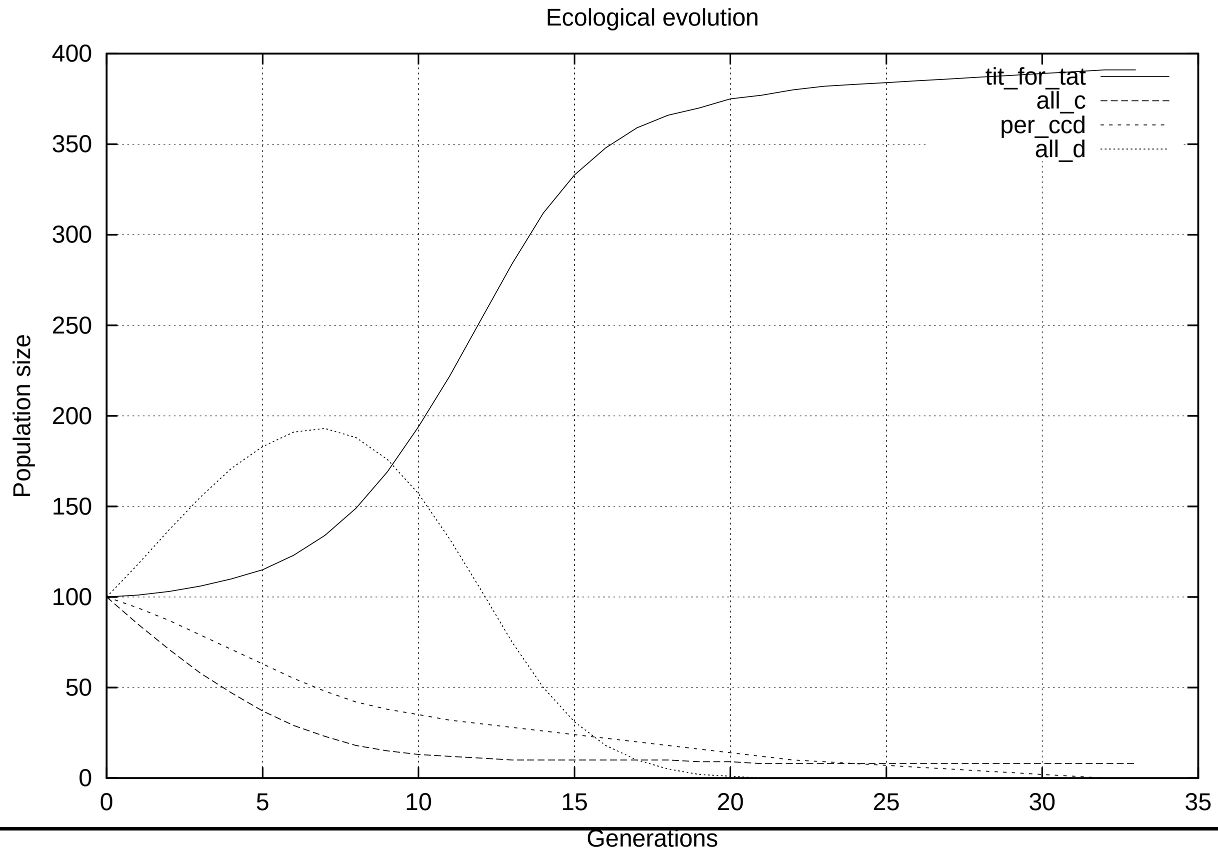
Exemples (tournoi)

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

$$\text{Classement} \left\{ \begin{array}{ll} 1 & \texttt{all\_d} \\ 2 & \texttt{tit\_for\_tat} \\ 3 & \texttt{per\_ccd} \\ 4 & \texttt{all\_c} \end{array} \right.$$

Exemples (évolution)

	1	2	3	4	...
all_c	100	85	71	58	...
all_d	100	118	137	155	...
per_ccd	100	94	187	79	...
tit_for_tat	100	101	103	106	...



Résultats classiques

– Résultats mis en avant par Axelrod

– Possibilité d'apparition d'un ordre sans autorité extérieure

– Critères de qualité pour une stratégie (en évolution) :

  – **bienveillance**

  – **réactivité**

  – **indulgence**

  – *simplicité (clarté du comportement)*

tit_for_tat

Au premier coup je coopère (C), ensuite si mon adversaire a coopéré (C) au coup précédent, je coopère (C), s'il a trahi (D), je trahis (D).

☞ tit_for_tat ne gagne jamais contre personne !

☞ Au mieux elle fait le même score.

☞ Mais, au pire elle ne perd que 5 points quel que soit l'adversaire et la longueur de la partie !

`gradual`

First opponent's defection : $\boxed{\texttt{D}} \longmapsto \boxed{\texttt{D}}\ \boxed{\texttt{C}}\ \boxed{\texttt{C}}$

Second opponent's defection : $\boxed{\texttt{D}} \longmapsto \boxed{\texttt{D}}\ \boxed{\texttt{D}}\ \boxed{\texttt{C}}\ \boxed{\texttt{C}}$

Third opponent's defection : $\boxed{\texttt{D}} \longmapsto \boxed{\texttt{D}}\ \boxed{\texttt{D}}\ \boxed{\texttt{D}}\ \boxed{\texttt{C}}\ \boxed{\texttt{C}}$

$$\vdots$$

`gradual` cooperates on the first move, then after the first opponent's defection defects once, and cooperates twice, after the second opponent's defection defects twice and cooperates twice, ..., after the n^{th} opponent's defection defects n times and cooperates twice.

bad_bet

Cooperates until an opponent's defection. As soon as its opponent has betrayed plays as follow :

1. During 4 moves it plays as `tit_for_tat`
2. then it plays as `all_c` during 4 moves
3. then it plays as `spiteful` during 4 moves
4. finally it plays as `per_ccd` during 4 moves
5. it then compares relative payoff limited to training period for each of those 4 strategies and chooses to play the most interesting one (the one which has given the highest payoff) during next 4 moves.
6. it updates choosen strategy's relative payoff
7. it then goes back to 5

Quelques apports de l'approche computationnelle

Deux gros problèmes :

- ☞ Comment automatiser la recherche d'une stratégie
- ☞ Comment évaluer *l*e plus objectivement possible une stratégie

Les solutions

- ☞ Simulations automatisables et reproductibles
- ☞ Comportement codable de manière infini

Pour l'évaluation d'une stratégie
- *absolue* : le rang dans une simulation avec une large population de stratégie
- *relative* : comparaison des rangs de 2 stratégies dans une simulation avec une large population de stratégies

Définir une classe de stratégies

- Pour automatiser la recherche il faut déterminer une méthode descriptive de définir un ensemble de stratégies
- Pour définir un ensemble de stratégies on peut par exemple :
  - définir une structure capable d'être décodée en un comportement à adopter face à un adversaire
  - utiliser toutes les manières possibles de remplir cette structure comme autant de stratégies

  ➠ approche génétique de définir des individus

Méthodes exhaustives risquent :
- de ne pas être objective
- jamais complètes
- inutilisable en retour (incompréhension des traits de la stratégie)

Considérons, par exemple, toutes les stratégies qui ne peuvent voir que leur dernier coup et le dernier coup de leur adversaire. Une de ces stratégies peut être définie par :

– au premier coup je joue $\boxed{\texttt{C}}$ puis

– $\left\{\begin{array}{l} \text{si j'ai joué } \texttt{C} \text{ et qu'il a joué } \texttt{C} \text{ alors je joue } \boxed{\texttt{C}} \\ \text{si j'ai joué } \texttt{C} \text{ et qu'il a joué } \texttt{D} \text{ alors je joue } \boxed{\texttt{D}} \\ \text{si j'ai joué } \texttt{D} \text{ et qu'il a joué } \texttt{C} \text{ alors je joue } \boxed{\texttt{D}} \\ \text{si j'ai joué } \texttt{D} \text{ et qu'il a joué } \texttt{D} \text{ alors je joue } \boxed{\texttt{D}} \end{array}\right.$

Le *génotype* de cette stratégie peut être noté comme

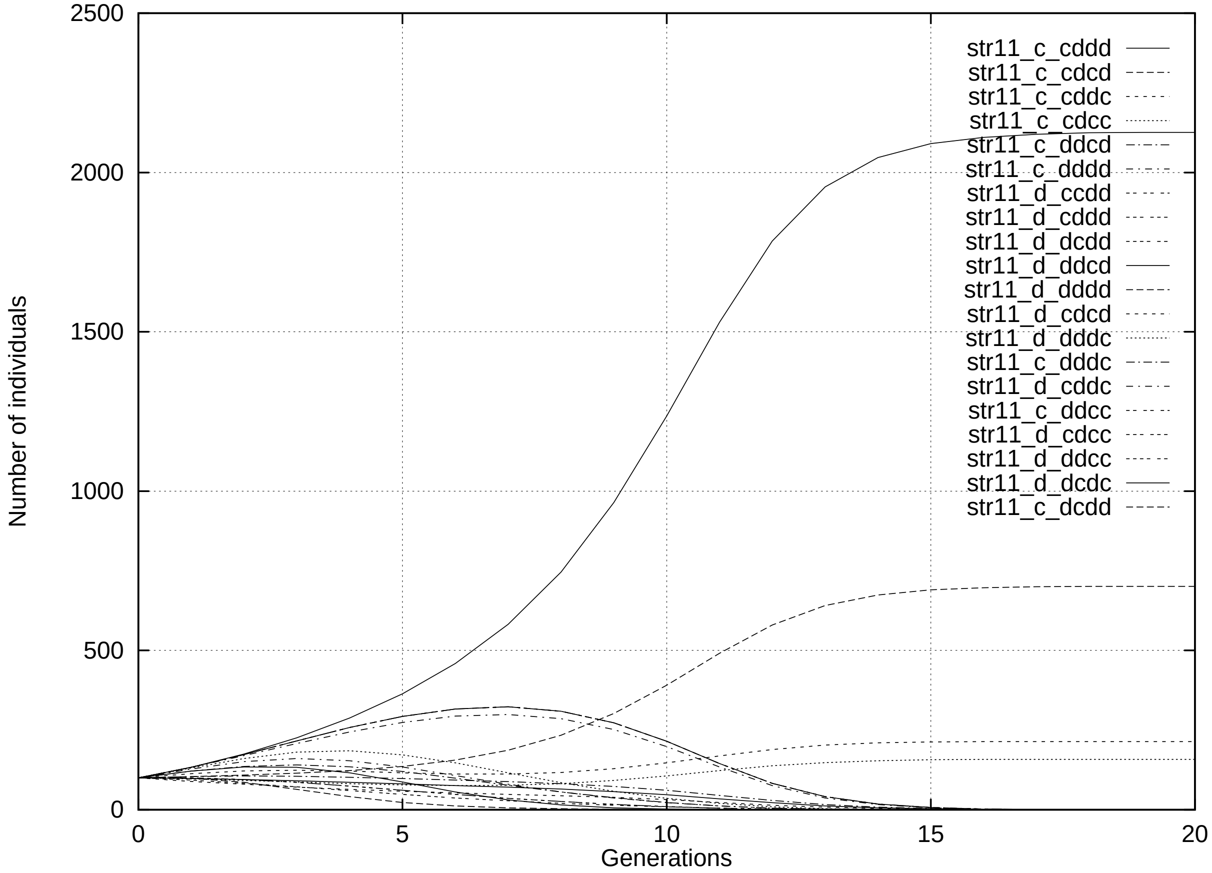
C	C	D	D	D

Avec ce génotype $2^5 = 32$ stratégies (incluant les classiques) sont définies.

Les classes `Memory`

- les classes `memory` :
  - Chaque stratégie voit seulement M coups de son passé, et O coups du passé de son adversaire
  - Une stratégie commence par jouer $\max(M, O)$ coups prédéfinis
  - La réponse à utiliser contre l'adversaire dépend uniquement de l'historique visible du jeu
  - $2^{\max(M,O)+2^{(M+O)}}$ stratégies décrites
- les classes `binary_memory` :
  - Mêmes idées que pour les classes `memory` avec ajout d'un indicateur de telle manière que la réponse à l'adversaire dépende :
    1. du nombre de trahisons de l'adversaire
    2. de l'historique visible du jeu
  - $2^{\max(M,O)+2^{(M+O+1)}}$ stratégies décrites

Un exemple d'évolution de classes complète



Quelques résultats

Nous avons évalués quelques stratégies en les ajoutant dans les évolutions de classes complètes. L'évaluation de la stratégie étant son rang à la fin de l'évolution. Voici quelques résultats :

	M	O	taille de classe	évaluation			
				`bad_bet`	`gradual`	`tit_for_tat`	`spiteful`
`memory`	0	1	8	1	1	1	1
	0	2	64	1	5	2	21
	1	1	32	1	2	3	1
	1	2	1024	1	6	13	37

Chaque évaluation nécessite de remplir une matrice *class size*$\times$*class size*

Ces expériences nous ont permis de découvrir de nouvelles bonnes stratégies, comme la gagnante de la classe `memory` $(M = 1,\, O = 2)$.

Partie C

Présentation du Projet

LIFL