Formation de coalitions répétée dans un contexte stochastique : protocoles et expérimentations

Josselin Gueneron

Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité INFORMATIQUE

Préparée au sein de l'Université de Caen Normandie

Formation de coalitions répétée dans un contexte stochastique : protocoles et expérimentations

Présentée et soutenue par
JOSSELIN GUENERON

Thèse soutenue le 13/12/2022 devant le jury composé de		
M. SAMIR AKNINE	Professeur des universités, Université Lyon 1 Claude Bernard	Rapporteur du jury
M. STÉPHANE AIRIAU	Maître de conférences, UNIVERSITE PARIS DAUPHINE	Membre du jury
MME AMAL EL FALLAH SEGHROUCHNI	Professeur des universités, Sorbonne Université	Membre du jury
M. MAXIME MORGE	Maître de conférences, Université de Lille	Membre du jury
M. RENE MANDIAU	Professeur des universités, Université polytechnique Hauts de France	Président du jury

Thèse dirigée par GREGORY BONNET (Groupe de recherche en informatique, image, automatique et instrumentation)

UNIVERSITÉ CAEN NORMANDIE

GREYC
Laboratoire de recherche en sciences du numérique

Remerciements

Ces trois années de thèse ont été pour moi une expérience incroyable d'une très grande intensité. Cela a notamment été pour moi l'occasion de faire bon nombre de rencontres lors de divers événements, qui ont donné lieu à des conversations très intéressantes.

Je vais commencer par remercier mon premier enseignant d'informatique, Ronan Charpentier, qui m'a donné l'envie et la motivation de continuer dans cette voie à l'université. Merci également à l'ensemble de mes enseignant.e.s de l'Université de Caen Normandie, et en particulier Étienne Grandjean, Jerzy Karczmarczuk et Emmanuel Cagniot.

Je tiens à remercier René Mandiau et Samir Aknine d'avoir accepté d'être les rapporteurs de ma thèse, ainsi que Stéphane Airiau, Amal El Fallah Seghrouchni et Maxime Morge d'avoir accepté de faire partie de mon jury de thèse. Merci également à Patrice Boizumault pour son implication dans mon comité de suivi individuel de thèse. Un immense merci à Grégory Bonnet, mon directeur de thèse, pour son implication, sa patience, ses encouragements, qui m'ont permis de réaliser cette thèse dans les meilleures conditions. J'ai grandement apprécié travailler avec lui et espère pouvoir continuer par la suite. Je tiens également à remercier l'ensemble de l'équipe MAD qui me supporte depuis ma 3ème année de licence, et qui m'a accueilli à bras ouverts lors de mes deux stages puis de cette thèse. C'est une équipe formidable avec une cohésion incroyable. Également merci à toutes les personnes qui contribuent à la bonne ambiance du laboratoire, et tout spécialement aux membres du groupe LGDLSPB, à savoir Pierre, Nadjet, Justine, Lauréline, Anaëlle, Gaétan, Céline, Matthieu, Alexis, Sébastien et Virginie. Merci à cette dernière ainsi qu'à Marie de l'école doctorale, avec qui les tâches administratives deviennent supportables, grâce à leur professionnalisme et leur grande efficacité.

Merci à ceux qui ont partagé avec moi le bureau S3-368, nos discussions, tant sérieuses que décontractées, vont me manquer. Je remercie donc chaleureusement Christopher, Sergej, Romain, Mihail, et plus spécialement Sébastien.

Ce dernier fait partie de mes amis proches, que je tiens particlièrement à remercier pour leur présence et leur amitié, avec qui je passe des moments inoubliables : Sébastien, Julien dit Pumba, Morgane, Louis-David, Rémi, Joscelyn, Clément, Marine, Jean-Philippe et Thibaut. J'adresse également toute ma reconnaissance à Varpu, ma meilleure amie, qui

malgré la distance est toujours là.

Une petite mention spéciale à Djé, Nanis, Ségnin, Gudj, Poulpy, Zack et Tessa, anciens et actuels membres du staff du Warpzone, où j'ai passé de nombreuses soirées très agréables.

Un grand merci à mes parents ainsi qu'à ma famille, qui m'ont toujours soutenu et encouragé, et tout fait pour que j'aille le plus loin possible. Des remerciements spéciaux à Clémentine, qui partage ma vie et me supporte depuis une demie décennie désormais, et à Loki pour ses miaulements inspirants.

Mes excuses à celles et ceux dont le nom n'apparaît pas ci-dessus, et je vous prie d'accepter mes remerciements. Merci à toutes et à tous.

TABLE DES MATIÈRES

INTRODUCTION

Contexte

Le domaine de l'intelligence artificielle cherche à proposer des solutions efficaces à des problèmes difficilement résolubles par un être humain. Dans un certain nombre de cas, cela passe par la conception d'entités autonomes douées de capacités spécifiques, appelées *agents artificiels*. Lorsque des agents existent dans un même environnement, nous parlons alors de *système multi-agents*. Ces derniers permettent de modéliser de nombreuses applications réelles (par exemple des chaînes logistiques ou des réseaux de capteurs), dans lesquelles les *agents* peuvent parfois être amenés à devoir agir ensemble pour réaliser un but commun, c'est-à-dire *coopérer*. Toutefois, certaines applications peut mettre en interaction des agents *égoïstes*, qui malgré le besoin de coopération, vont souhaiter en retirer un profit maximal. Dans un tel cadre, les agents doivent donc trouver un équilibre entre leur profit personnel et la coopération à l'échelle globale.

La *formation de coalitions* est un problème de la théorie des jeux permettant de modéliser des systèmes multi-agents où les agents doivent former des *coalitions* afin d'effectuer leurs tâches. C'est donc un cadre privilégié pour modéliser les problèmes de coopération. Toutefois, ce cadre fait couramment des hypothèses qui rendent difficile la modélisation d'applications réelles. En effet, lorsque des agents coopèrent, ils ne connaissent pas nécessairement les compétences des autres, c'est-à-dire ils ne possèdent pas de *connaissance a priori*. Or, le fait que les agents possèdent une connaissance *a priori* des compétences et utilités des autres agents est une première hypothèse courante de la formation de coalitions. Une autre hypothèse est celle de *déterminisme* : un même groupe d'agents aura toujours la même efficacité. Or, ceci ne permet pas de modéliser le fait qu'un agent puisse être soumis à des aléas, soit intrinsèques, soit environnementaux. Enfin, une dernière hypothèse décrit le fait que la décision de la formation des différentes coalitions se fait de manière *centralisée*, c'est-à-dire qu'une seule et même entité prend la décision pour l'ensemble des agents. Cependant, dans nombre d'applications réelles, cette hypothèse n'est pas compatible avec les contraintes spatiales ou de ressources des agents.

Nous proposons donc dans ce manuscrit d'étudier la levée de ces différentes hypo-

thèses imcompatibles avec nombre d'applications réelles en proposant des modèles et des protocoles, analysés par expérimentations et une étude empirique des résultats.

Contributions

Notre étude sur la levée des différentes hypothèses se décompose en deux parties. Nous aborderons dans un premier temps la levée des hypothèses de déterminisme et de connaissance *a priori* des utilités, par la proposition d'un modèle pour la formation de coalitions stochastique répétée qui s'abstrait de ces hypothèses ainsi que deux concepts de solutions fondés sur une notion d'équilibre exploration-exploitation bien connue du domaine de l'apprentissage par renforcement.

Afin d'aborder la problématique concernant la décentralisation du problème de formation de coalitions, nous proposerons l'adaptation d'un protocole de concessions monotones pour la négociation multilatérale entre agents utilitaristes, proposé par Ulle Endriss. En effet, ce protocole est exempt d'hypothèse sur la structure du système et ne contient aucune entité centrale telle qu'un commissaire-priseur, qui est un élément simulant une centralisation du système. Nous proposerons donc dans un premier temps une adaptation de ce protocole à la formation de coalitions dans un cadre distribué, avant d'en proposer une version dans un cadre décentralisé. Enfin, nous travaillerons à l'unification de l'ensemble des modèles créés pour la levée des différentes hypothèses, en proposant une adaptation du protocole distribué à un cadre stochastique et répété, puis une adaptation dans ce même cadre du protocole décentralisé. Ainsi, les trois hypothèses dont nous souhaitons nous abtraire seront traitées dans un même et unique modèle.

Organisation du document

Ce manuscrit est organisé en six chapitres, plus une conclusion : deux chapitres d'état de l'art, deux chapitres de contributions pour les jeux de coalitions stochastiques répétés, et enfin deux chapitres de contributions concernant l'adaptation d'un protocole de concessions monotones à la formation de coalitions.

Les chapitres 1 et 2 dressent donc un état de l'art. Dans un premier temps, dans le chapitre 1, nous présentons les notions que nous utiliserons dans la suite concernant les propriétés des agents et systèmes multi-agents, ainsi que les concepts fondamentaux de la théorie des jeux coopératifs, avec une focalisation sur le cadre de la formation de

coalitions à utilité transférable. Nous y présentons ainsi certaines spécificités de ce cadre telles que les concepts de solutions, les modèles intégrant des contraintes particulières, ainsi que les méthodes de résolution. Le chapitre 2 quant à lui nous permet de présenter les travaux de la littérature qui traitent des questions connexes aux nôtres, à savoir la levée des hypothèses de déterminisme et de connaissance *a priori* des utilités des coalitions. Ce chapitre présente également un problème d'apprentissage par renforcement, à savoir le problème des bandits manchots, avec lequel nous faisons dans ce manuscrit une analogie à la formation de coalitions stochastique répétée.

Le chapitre 3 est consacré à la proposition d'un modèle de jeux de coalitions stochastiques répétés ainsi qu'à la proposition de deux concepts de solutions fondés sur un équilibre exploration-exploitation inspiré des stratégies du problème des bandits manchots. Le chapitre 4 est dédié à la mise en place d'un cadre d'expérimentation pour ces contributions, avec deux méthodes d'apprentissage différentes.

Dans le chapitre 5, nous proposons une adaptation d'un protocole de concessions monotones pour la négociation mutlilatérale au cadre de la formation de coalitions. Nous y proposons notamment de nouvelles stratégies pour notre protocole adapté, avant de modifier ce dernier afin de pouvoir l'appliquer à un cadre décentralisé, ici encore avec de nouvelles stratégies. Le chapitre 6 est dédié à l'extension du protocole adapté à la formation de coalitions à un cadre stochastique, en s'appuyant sur les travaux menés dans les chapitres 3 et 4, notamment en réutilisant l'analogie avec les bandits manchots afin de proposer des stratégies intégrant une notion d'exploration. Enfin, nous proposons dans ce chapitre une adaptation du protocole stochastique au cadre décentralisé, afin de proposer un modèle et des stratégies répondant à l'ensemble de nos problématique.

Nous clôturons ce manuscrit par un chapitre présentant une synthèse de nos travaux ainsi que des perspectives pour de futurs travaux, par exemple des pistes afin d'affiner la distribution des gains dans les divers protocoles proposés.

Chapitre 1

Former des coalitions pour coopérer

Sommaire

Les systèmes multi-agents (SMA) permettent de modéliser un grand nombre de cadres applicatifs réels, comme des réseaux de capteurs [Glinton *et al.*, 2008], des chaînes logistiques [Gaston et desJardins, 2005, Adam *et al.*, 2011], des modèles de patrouille [Othmani-Guibourg *et al.*, 2017], des constellations de satellites [Bonnet et Tessier, 2007], ou encore des réseaux électriques intelligents [Bremer et Lehnhoff, 2017]. Dans ce chapitre, nous verrons tout d'abord comment les agents peuvent s'articuler en groupes, nommés coalitions, afin de réaliser des tâches conjointement, à quoi sert cette coopération, puis des aspects plus spécifiques de la formation de coalitions dans le cadre d'utilité transférable, c'est-à-dire où les agents doivent répartir l'utilité de chaque coalition entre ses membres. Nous prendrons en exemple fil rouge un problème concret qui est la gestion d'un port maritime.

Dans un tel contexte, les agents sont très nombreux (opérateurs, navires, douaniers...), et les objectifs peuvent être multiples.

Exemple 1. *Les opérateurs ont en charge la fluidification du trafic du port avec notamment la gestion du flux des navires dans la digue, tandis que les dockers doivent s'occuper du chargement et déchargement des conteneurs. Ces mêmes conteneurs doivent également être inspectés par les douaniers avant que les transporteurs routiers prennent en charge la marchandise.*

La modélisation de tels cadres applicatifs grâce aux systèmes multi-agents peut avoir divers objectifs. Par exemple, cela peut consister en la recommandation d'actions pour les différents acteurs sur le port, de la planification ou encore en la formation de groupes pour certaines tâches (associer un groupe de navires aux quais, les déchargeurs aux navires, les douanes aux conteneurs, etc...). Un élément important pour la modélisation de tels problèmes est l'évaluation de l'intérêt que les agents ont à exécuter leurs actions, ce qui est appelé *utilité*. Cette utilité est définie par une *fonction objectif* qui caractérise le but de l'agent. Les agents sont alors considérés comme *utilitaristes*.

Exemple 2. *L'utilité d'un navire est dépendante du temps passé dans le port. Celle des dockers – qui chargent et déchargent les navires – est liée au nombre de conteneurs déplacés.*

Cependant, l'environnement de cadres applicatifs comme un port maritime est souvent dynamique et soumis à des aléas comme la météo, ou juste des difficultés de terrains. Cela peut rendre difficile la modélisation des fonctions objectif, par exemple en modifiant l'utilité associée aux diverses entités en raison du dynamisme et des aléas. Ces problèmes sont donc à prendre en compte et nous présentons dans la suite des approches pour cela.

1.1 Propriétés et coopération

Les systèmes multi-agents, tout comme les agents qui les composent, peuvent posséder des propriétés intéressantes pour modéliser de telles applications réelles. Cette section a pour but de présenter les propriétés qui nous seront utiles par la suite, concernant la nature utilitariste des agents. Cette caractérisation des agents nous amènera à la question de la coopération au sein de ces systèmes multi-agents, dans un but de maximisation de l'utilité générée. Ce faisant, nous nous intéresserons à certaines questions de modélisation

comme la gestion de l'incertitude et des connaissances imparfaites dans les modèles afin que les agents puissent malgré tout effectuer des actions efficaces et coopérer dans des conditions favorables pour eux.

1.1.1 Des agents utilitaristes

Un agent est une entité du système, au sein duquel il peut évoluer de différentes façons, comme par exemple faire une action spécifique, obtenir des gains, ou encore simplement observer son environnement. Un agent peut aussi posséder des propriétés qui caractérisent en partie sa nature. Il existe beaucoup de propriétés, mais intéressons-nous ici à des propriétés spécifiques qui caractérisent un comportement utilitariste, c'est-à-dire pour lesquels la notion d'utilité est centrale, comme pour les agents de notre exemple de port maritime. L'utilité ne doit pas être confondue avec le gain. Alors, que ce dernier caractérise ce qui est acquis suite à une action par exemple, l'utilité regroupe l'ensemble des éléments à prendre en compte afin de prendre la meilleure décision [Paccagnan *et al.*, 2022].

1.1.1.1 Rationalité

La propriété de rationalité d'un agent est la caractérisation du fait qu'il cherchera à maximiser son utilité [Morgenstern et Von Neumann, 1953, Simon, 1969].

Exemple 3. *Dans le contexte du port maritime, un opérateur cherchera à maximiser le nombre de navires qui peuvent s'amarrer ou le nombre de conteneurs déplacés.*

Cela peut donc concerner des aspects quantitatifs, mais également des aspects qualitatifs comme la satisfaction des clients (par exemple, la compagnie X sera « satisfaite » ou « non satisfaite »). Dans ce cas, la rationalité sera de maximiser cette satisfaction. Si les agents évoluent dans un environnement dynamique – c'est-à-dire qui change au cours du temps – alors leur rationalité peut être *limitée*. La rationalité limitée caractérise le fait qu'un agent cherchera à maximiser son utilité de manière *satisfaisante* dans le temps, plutôt que de chercher une maximisation immédiate [Russell et Norvig, 1995, Simon, 1969].

Exemple 4. *Un opérateur de port maritime préférera donner l'autorisation de s'amarrer à deux navires à la première heure, puis trois navires plus tard, plutôt que de donner l'autorisation à quatre navires dès la première heure car cela pourrait risquer de bloquer le port par la suite.*

1.1.1.2 Égoïsme

L'égoïsme est une propriété qui caractérise le fait que l'agent définit son utilité uniquement par rapport à son gain personnel [Shehory, 1998]. Il aura donc un comportement individualiste en toutes circonstances, quitte à être pénalisé sur d'autres aspects. Tout comme pour la rationalité, l'égoïsme peut également concerner des aspects quantitatifs (gains, ressources, etc.) ou qualitatifs (niveau de satisfaction personnelle par exemple).

Exemple 5. *Un exemple d'agent égoïste peut également être illustré par les compagnies passant par le port maritime, qui souhaitent être prises en charge le plus vite possible, et ne tiennent pas compte des autres compagnies.*

De façon générale, la notion d'égoïsme est opposée à celle de l'altruisme. Par la suite, les agents auxquels nous nous intéresserons seront, sauf indication contraire, des agents rationnels et égoïstes.

1.1.2 Propriétés d'architecture des systèmes multi-agents

Tout comme les agents, les systèmes multi-agents peuvent posséder des propriétés caractérisant certains aspects les concernant. Un de ces aspects le plus courant est l'architecture de ces systèmes qui régit les interactions entre les agents. Nous parlons alors de centralisation ou non du système.

1.1.2.1 Du système centralisé au système décentralisé

Nous appellons système centralisé tout système au sein duquel l'ensemble des calculs est exécuté par une seule entité, qu'elle soit ou non un des agents du système. Tous les agents partagent les mêmes connaissances sur leur environnement, et les observations effectuées par les agents sont communes [Vercouter, 2000].

Exemple 6. *Une équipe de dockers sous le contrôle d'un chef d'équipe, où le chef décidera quels conteneurs décharger et indique à chaque docker de son équipe le conteneur dont il devra s'occuper, est un système centralisé. Aucun docker n'exécutera une tâche sans l'aval de son chef d'équipe, et il lui rapporte toute information.*

Beaucoup de travaux dans le domaine des systèmes multi-agents portent sur la distribution du système [Moulin et Chaib-Draa, 1996]. L'intérêt principal de cette distribution est de paralléliser les calculs à effectuer entre les agents, afin de profiter des ressources

de chacun. Chaque agent peut alors résoudre une partie du problème, suivi de la mise en commun des résultats distribués. Ensuite, une entité centrale agrège les résultats, prend une décision selon ces derniers, puis détermine les actions que chaque agent désormais doit exécuter. Également, dans les systèmes distribués, si des informations peuvent être connues des agents, alors celles-ci sont généralement communes à tous.

Exemple 7. *Reprenons l'exemple précédent mais supposons maintenant que l'équipe de dockers ne possède pas de chef d'équipe. Les dockers vont pouvoir observer quels conteneurs doivent être déchargés, échanger l'information avec leurs coéquipiers, et se répartir les conteneurs à décharger entre eux. Ainsi, chaque docker peut par exemple se proposer pour le déchargement d'un conteneur, et les autres seront au courant. Ceci est un système distribué.*

Les systèmes décentralisés sont quant à eux des systèmes où toute centralisation a disparu. Il n'y a plus de mise en commun des calculs, chaque agent calcule ce qui lui est utile de son côté, et il n'y a aucune décision centralisée [Brooks, 1986]. Si les agents peuvent avoir une base de connaissances commune dans les systèmes centralisés et distribués, il n'en est rien dans les systèmes décentralisés. En effet, les agents possèdent chacun leurs connaissances propres, qu'ils peuvent affiner par le biais des observations qu'ils peuvent faire. Des échanges d'informations entre agents sont néanmoins possibles, mais avec un coût de communication et de ressources en général.

Exemple 8. *Notre équipe de docker devient un système décentralisé si chaque docker, toujours sans chef d'équipe, va observer les conteneurs à décharger, en choisir un et aller le décharger sans partager cette information à ses coéquipiers. Dans ce cadre, certains dockers peuvent donc savoir qu'un certain conteneur est déjà déchargé tandis que d'autres ne le sauront qu'en l'observant eux-mêmes.*

La distribution et la décentralisation d'un système permettent en général un gain de temps, en raison de la parallélisation des calculs. Cependant, les solutions calculées de cette manière ne sont pas nécessairement optimales en raison de la localité de l'exécution des calculs.

1.1.2.2 Ouverture d'un système

Les sytèmes multi-agents peuvent être ouverts ou fermés. Ce qui définit l'ouverture d'un système est la capacité qu'ont les agents à entrer ou sortir de ce système à tout moment [Hewitt, 1991]. Par exemple, cela peut être un agent défectueux ou non-désireux de

continuer à faire partie du système qui en sort, ou bien au contraire, un agent de remplacement ou tout simplement nouveau qui souhaite y rentrer. Cette propriété permet une certaine flexibilité dans les systèmes : dans notre exemple de port, cela permet l'ajout et la suppression dynamique de compagnies transitant par le port. En revanche, la contrepartie à cette ouverture est la complexité qu'elle amène, et parfois même des risques [Vercouter, 2000]. D'une part, la complexité est augmentée car il faut pouvoir gérer les entrées et les sorties du système, comme par exemple ajouter dynamiquement des informations aux connaissances des agents, ou redistribuer des ressources ou tâches à réaliser. D'autre part, l'ouverture d'un système peut parfois ouvrir la voie aux manipulations. L'exemple du port maritime peut à nouveau illustrer ceci. Imaginons une compagnie transitant par le port, et que pour une raison quelconque, celle-ci n'y soit plus admise. Une manipulation pourrait consister à simplement revenir sous une nouvelle identité afin de contourner l'interdit.

Les propriétés d'ouverture et de centralisation du système spécifient des contraintes sur l'architecture des systèmes multi-agents, et bien que cela ne change pas la caractérisation des agents, la structure du système dans lequel ils évoluent peut avoir une influence sur leurs choix. En effet, selon le type d'architecture, l'exécution des tâches, les connaissances et les observations ne sont pas communes. Si nous reprenons nos agents rationnels et égoïstes (pour qui le gain personnel est le principal intérêt donc), ceux-ci peuvent donc rencontrer des difficultés à obtenir un gain personnel élevé (par exemple en raison du manque d'observations), ou même dans certains cas, ne seront pas en mesure d'exécuter leurs tâches seuls (par exemple pour une tâche trop complexe ou dû à un manque de connaissances). C'est alors que la coopération entre les agents peut devenir intéressante pour eux.

1.1.3 Interaction des agents utilitaristes

Dans certains cadres applicatifs des systèmes multi-agents, la question de la coopération peut se poser. Telle que décrite par Morgenstern et Von Neumann [Morgenstern et Von Neumann, 1953], la coopération entre agents consiste en la formation de groupes nommés *coalitions*. Cependant, avant d'en arriver là, il nous faut nous poser certaines questions : quel cadre pour modéliser ces interactions ? Qu'auraient à gagner les agents s'ils coopèrent ? De quelle manière peuvent-ils décider avec qui coopérer ? Nous pouvons trouver des réponses à ces questions dans la *théorie des jeux*.

1.1.3.1 Que permet la théorie des jeux ?

La théorie des jeux, dont les fondements ont été proposées par Morgenstern et Von Neumann [Morgenstern et Von Neumann, 1953], est un domaine de recherche portant sur les interactions entre agents et les stratégies que ces derniers doivent adopter selon la situation. Un *jeu* caractérise simplement une situation comprenant des agents.

Définition 1.1 (Jeu)**.** *Un jeu $\mathcal{G}$ de la théorie des jeux est un tuple comprenant au moins un ensemble N de* joueurs *(que nous appelons également agents) avec $|N| \geq 2$.*

La théorie des jeux regroupe différents types de jeux, selon la situation à représenter. Voici une liste non-exhaustive de types de jeux parmi les plus courants, ainsi qu'une description informelle des cadres associés.

- Jeux à somme nulle : ce sont des jeux strictement compétitifs où les gains et les pertes des joueurs s'équilibrent et dont la somme est égale à 0. Autrement dit, dans un jeu mettant en scène deux joueurs, le gain d'un joueur correspond à la perte de l'autre joueur. Des exemples courants sont le poker ou encore le tarot, respectivement où les sommes des gains/pertes et des points sont égales à 0.
- Jeux simultanés : dans ces jeux, les agents doivent décider en même temps d'une action à entreprendre. Cela est très bien illustré par le jeu du pierre-papier-ciseaux, où chaque joueur décide de jouer un des trois objets (pierre, papier ou ciseaux donc) en même temps que son adversaire, et si un joueur joue un objet qui domine l'autre (exemple : la pierre domine les ciseaux) alors il gagne un point. Lors d'une égalité, aucun joueur ne marque de point.
- Jeux séquentiels : ils s'opposent par définition aux jeux simultanés. Ici, les joueurs vont devoir décider de l'action à exécuter à tour de rôle, et notamment en prenant en compte ce qui a été fait par les autres joueurs avant cela. Il y a donc une notion de dynamique dans le jeu et les gains des joueurs (dépendant de l'historique du jeu) ne sont attribués qu'à la fin. Des exemples bien connus de tels jeux sont le jeu d'Échecs et le jeu de Go, où les joueurs jouent alternativement des pièces (blanches et noires), et où les actions des joueurs sont conditionnées par ce que leur adversaire a précédemment joué.
- Jeux coopératifs : ces jeux permettent de modéliser une situation où les joueurs ont la possibilité de coopérer ensemble, et ce même dans un contexte compétitif, par exemple pour réaliser une tâche trop difficile à exécuter individuellement. Un exemple, tiré de jeux de rôle en ligne massivement multijoueurs tels que World of

Warcraft, est lorsque des joueurs vont former des groupes pour réaliser des quêtes ensemble, comme pour vaincre un ennemi trop puissant pour être défait à un seul joueur.

Le but des agents au sein d'un jeu est donc d'adopter une stratégie qui, d'une part, maximisera leur propre gain (car ils sont égoïstes), et d'autre part, n'aidera pas les joueurs opposants (car ils sont rationnels et cela conduit à la dégradation de leur propre gain).

Nous pouvons d'ores et déjà identifier certains de ces jeux comme des jeux dits d'opposition (ou non-coopératifs), c'est-à-dire strictement compétitifs : les jeux à somme nulle et séquentiels. Notre intérêt résidant dans la modélisation de systèmes tels qu'un port maritime, nous pouvons exclure la possibilité de modéliser les interactions des agents par des jeux à somme nulle et des jeux séquentiels car maintenir une somme des gains et des pertes nulle ou que les agents agissent un par un ne représenterait pas convenablement les réalités de terrain, et ce en raison de la nature strictement compétitive de ces deux types de jeu. En effet, dans un tel cadre, des tâches peuvent être à réaliser collectivement. Prenons par exemple des douaniers devant inspecter des conteneurs : ces derniers peuvent très bien réaliser leur tâche, et ce sans compétition ni collaboration. Cependant, cela amène d'autres questions : est-ce que les douaniers seraient plus efficaces à inspecter un seul conteneur ensemble, puis passer au suivant, ou bien chacun inspecte un conteneur distinct ? Un autre exemple est si ces tâches demandent des ressources. Si une tâche simple apporte peu de gain et qu'une tâche complexe apporte un gain important, alors un agent égoïste aura de quoi hésiter. En effet, d'un côté, il souhaitera dépenser le moins de ressources possibles, mais son gain sera assurément faible. Tandis qu'en coopérant avec d'autres agents, ils pourraient collectivement remporter un gain plus important. Un tel exemple peut se retrouver également dans les dockers qui chargent ou déchargent les conteneurs, dont les ressources sont l'énergie et le temps.

En résumé, si certains jeux ne permettent pas la coopération, ne pas coopérer lorsque c'est possible peut paraître sous-optimal, car chacun cherchera à saborder les autres pour maximiser leur propre profit, ce qui peut, comme nous le verrons ci-dessous en s'appuyant sur les jeux simultanés, être néfaste pour tous les agents.

1.1.3.2 Compétition contre coopération

Si dans certains contextes la coopération n'est pas possible, dans d'autres elle peut apporter aux agents une plus-value, comme par exemple davantage de gains. Par exemple, prenons deux compagnies devant négocier un contrat pour le transports de conteneurs :

chaque compagnie peut soit décider de partager le transport avec l'autre, soit demander à avoir l'exclusivité du contrat. Si les deux décident de demander l'exclusivité, le propriétaire des conteneurs peut décider de prendre aucune des deux compagnies, alors celles-ci se retrouvent dans la pire situation.

Nous pouvons abstraire cet exemple via le dilemme du prisonnier [Flood *et al.*, 1950, Poundstone, 1993]. Ce dilemme, modélisé par un *jeu simultané*, met en scène deux agents complices d'un crime interrogés par la police à propos de ce dernier. Ils ne peuvent pas communiquer entre eux, et doivent choisir soit de se taire, soit de dénoncer l'autre agent. Si les deux agents décident de se taire, alors ils écoperont tous les deux d'une petite peine de prison. Si un des deux agents décide de se taire et l'autre de dénoncer son complice, alors celui qui s'est tu ira en prison, tandis que le délateur sera libre. Si les deux agents décident de dénoncer l'autre, alors les deux auront une lourde peine de prison. Une stratégie pour un tel jeu est donnée par l'équilibre de Nash [Nash, 1951, Osborne et Rubinstein, 1994]. Cet équilibre caractérise une stratégie à adopter pour maximiser son gain (ou minimiser son regret) en prenant en compte les choix des autres agents, lorsque l'ensemble des agents doivent agir de façon simultanée.

Si nous appliquons notre exemple précédent à un jeu de dilemme du prisonnier, alors la demande de l'exclusivité du contrat correspond à la dénonciation du partenaire, tandis que le partage du contrat correspond à la coopération (c'est-à-dire de ne pas dénoncer le complice).

Exemple 9. *Considérons la matrice de pertes (gains inversés) suivante :*

	exclusivité	*pas d'exclusivité*
exclusivité	*(3,3)*	*(5,0)*
pas d'exclusivité	*(0,5)*	*(0.5,0.5)*

Chaque agent – représentant chacun une des compagnies négociant le contrat – raisonne d'une façon locale, en tentant de prédire le choix de l'autre agent, et cherche donc ce qui serait le mieux pour lui compte tenu de ce choix. Leur raisonnement peut être résumé comme suit :

- *Si l'autre agent demande l'exclusivité, alors j'ai intérêt à la demander également, comme cela je perdrai 3 millions d'euros au lieu de 5 millions.*
- *Si l'autre agent ne demande pas d'exclusivité, j'ai également intérêt à demander l'exclusivité car je n'aurai aucune perte au lieu de perdre un demi million d'euros.*

La stratégie dominante découlant de ce raisonnement, donnée par l'équilibre de Nash, est donc que les deux agents demandent l'exclusivité, et perdent ainsi 3 millions d'euros chacun, tandis que s'ils avaient décidé tous les deux de partager le contrat, ils auraient perdu un demi million d'euros chacun au lieu des 3 millions comme en suivant leur stratégie.

Bien qu'étant une stratégie dominante, la non-coopération n'est donc pas une stratégie qui permet aux agents d'atteindre le meilleur résultat dans un tel cadre, sachant que la coopération aurait largement réduit leurs pertes. Cette conclusion est également celle apportée par Axelrod [Axelrod et Hamilton, 1981]. Ce dernier a étudié le dilemme du prisonnier itéré : le jeu est répété et les gains (ou pertes) sont mémorisés. Il a notamment montré que les stratégies égoïstes ne sont pas viables, tandis que les stratégies altruistes tendent à être récompensées. La coopération est donc également indiquée dans un cadre répété. Le dilemme du prisonnier peut également mettre en relation davantage que deux agents [Marwell et Schmitt, 1972], cependant, cela reste dans une configuration d'actions binaire telle que dénonce/ne dénonce pas. Ceci pose des limites lorsque le but est de modéliser des systèmes plus complexes où les actions peuvent être plus nombreuses, ou que les agents ne souhaitent pas nécessairement coopérer avec tout le monde. Pour ces raisons, nous allons nous concentrer sur le domaine de recherche au sein de la théorie des jeux qui s'intéresse à ces questions de coopération plus complexes : la *théorie des jeux coopératifs*.

1.2 Théorie des jeux coopératifs

Ce domaine de la théorie des jeux s'intéresse donc à la coopération dans les jeux. Cependant, notre intérêt étant porté sur les systèmes multi-agents, nous considérons le formalisme décrit par par Morgenstern et Von Neumann [Morgenstern et Von Neumann, 1953] où la coopération entre agents consiste en la *formation de coalitions*. En effet, la *formation de coalitions* est un problème permettant de modéliser la coopération des agents en groupes distincts – les *coalitions* – au sein du système, et ce sans contraintes fortes sur le nombre ou la nature des agents. La formation de coalitions peut par exemple permettre de former des groupes d'agents sur le port maritime, pour qu'ils effectuent des tâches ensemble, comme par exemple le déchargement de conteneurs.

Définition 1.2. *Soit $N = \{a_1, \ldots, a_n\}$ un ensemble d'agents. Une* coalition *C est un sous-ensemble de N tel que $C \subseteq N$. Lorsque la coalition C contient l'ensemble des agents*

de N, nous parlons de grande coalition*. Lorsqu'une coalition contient un seul agent, nous parlons de* coalition singleton*.*

Le formalisme de la formation de coalitions s'exprime à travers un *jeu de coalitions*. Le but de ces jeux est donc de former un ensemble de coalitions, appelé *structure de coalitions*. Dans les jeux de coalitions classiques, c'est-à-dire la forme décrite par Morgenstern et Von Neumann et la plus commune [Morgenstern et Von Neumann, 1953, Osborne et Rubinstein, 1994], les coalitions sont disjointes une à une, chaque agent ne faisant partie que d'une seule coalition. Une structure de coalitions forme donc une partition de N.

Définition 1.3 (Structure de coalitions)**.** *Une* structure de coalitions $\mathcal{CS}$ *est une partition de l'ensemble des agents de N en k coalitions disjointes : $\mathcal{CS} = \{C_1, \ldots, C_k\}$, $\bigcup_{C \in \mathcal{CS}} = N$ et $C_i \bigcap C_j = \emptyset,\ \forall i, j \in \{1, \ldots, k\}$ et $i \neq j$.*

Il existe deux grands types de ces jeux : ceux à utilité transférable, et leur généralisation, les jeux à utilité non-transférable. Les premiers permettent la répartition des utilités des coalitions entre leurs membres, tandis que dans les deuxièmes, le gain des agents est fixé au préalable.

1.2.1 Jeux à utilité non-transférable

Ce cadre sert à modéliser une formation de coalitions où, lorsque les coalitions sont formées, les agents reçoivent un certain gain fixé préalablement, sans possibilité de transferts de gains entre les agents. Les agents doivent donc préalablement déterminer quelles coalitions ils préférent afin d'en retirer le gain maximal car, rappelons-le, ils sont rationnels et égoïstes.

1.2.1.1 Jeux de coalitions quantitatifs et hédoniques

Une première forme de jeux de coalitions à utilité non-transférable sont les *jeux de coalitions quantitatifs* [Bonzon, 2007, Chalkiadakis *et al.*, 2011]. Un tel jeu est défini par un ensemble d'agents, et une fonction caractéristique qui attribue pour chaque coalition un vecteur de gain prédéfini, spécifiant le gain de chaque agent faisant partie de cette coalition.

Définition 1.4 (Jeu de coalitions quantitatif)**.** *Un* jeu de coalitions quantitatif *est un tuple $\mathcal{G} = \langle N, v \rangle$ où $N = \{a_1, \ldots, a_n\}$ est un ensemble d'agents, et $v : 2^N \rightarrow \mathbb{R}^C$ est*

une fonction caractéristique qui associe à chaque coalition $C \subseteq N$ *un vecteur* $\vec{v}(C) = \{x_1, \ldots, x_k\}$ *où* x_i *est le gain de l'agent* $a_i \in C$ *à la formation de la coalition* C*.*

Une solution à un jeu de coalitions quantitatif est donc une structure de coalitions $\mathcal{CS}$. Les agents doivent ainsi déterminer les coalitions qu'ils souhaiteraient former en priorité en fonction de leur intérêt. Une solution doit alors satisfaire un critère de *stabilité*, c'est-à-dire qu'aucun agent ne souhaite former une autre coalition – ce qui est appelé une *déviation* – que celle à laquelle il appartient dans la solution.

La forme la plus générale des jeux à utilité non-transférable sont les *jeux de coalitions hédoniques* [Chalkiadakis *et al.*, 2011, Vallée et Bonnet, 2017]. Ces jeux ne possèdent pas de fonction caractéristique mais utilisent un ensemble de préférences définies par les agents sur les coalitions. Ces jeux hédoniques permettent de modéliser des jeux plus spécifiques comme les jeux de coalitions quantitatifs.

Définition 1.5 (Jeu de coalitions hédoniques)**.** *Un jeu de coalitions hédoniques est un tuple* $HG = \langle N, \succeq \rangle$*, où* $N = \{a_1, \ldots, a_n\}$ *est l'ensemble des agents, et* $\succeq = \{\succeq_1, \ldots, \succeq_n\}$ *un ensemble de préférences des agents.*

Les jeux hédoniques intègrent donc une notion de *préférences* [Hansson, 1968], dont les agents usent pour ordonner les coalitions dont ils peuvent faire partie, de celles qu'ils souhaitent le plus former à celles qu'ils veulent le moins [Vallée et Bonnet, 2017]. Formellement, ces préférences sont définies comme suit.

Définition 1.6 (Préférences)**.** *Une relation de préférence* $\succeq_i \in \succeq$ *pour un agent* $a_i \in N$ *dans un jeu* $HG = \langle N, \succeq \rangle$ *est un préordre total avec indifférence sur l'ensemble* G_i *des coalitions auxquelles* a_i *peut appartenir, avec* G_i *défini comme suit :*

$$G_i = \{C \mid \forall C \in 2^N \text{ } t.q. \text{ } a_i \in C \}$$

Les agents utilisant ces relations de préférence sont, de par la nature de ces dernières, rationnels [Suzumura, 1976]. Les jeux hédoniques peuvent donc modéliser des jeux de coalitions quantitatifs car la fonction caractéristique de ces derniers peut être exprimée par des préférences. En effet, prenons deux coalitions C_1 et C_2 dont l'agent a_i peut faire partie. Si nous avons, dans un cadre de jeux de coalitions quantitatifs, $x_i(C_1) \geq x_i(C_2)$, cela peut être traduit en la relation $C_1 \succeq_i C_2$ dans un jeu hédonique [Shapley et Shubik, 1974].

1.2.1.2 Stabilité

Afin de déterminer quelle structure de coalitions $\mathcal{CS}$ former selon ces préférences, il existe des concepts de solutions, qui caractérisent une notion de *stabilité* particulière. La stabilité d'une solution est le fait qu'aucun agent ne souhaite ou ne peut dévier, c'est-à-dire quitter la coalition $C_i(\mathcal{CS})$ dans laquelle il est dans la solution $\mathcal{CS}$ pour en rejoindre une autre. Certains concepts de solutions pour les jeux de coalitions hédoniques sont présentés formellement dans la table 1.1 [Vallée et Bonnet, 2017].

Stabilité de Nash	$\forall a_i \in N, \nexists C \in \mathcal{CS} \cup \{\emptyset\} : C \ \cup \{a_i\} \succ_i C_i(\mathcal{CS})$
Stabilité individuelle	$\forall a_i \in N, \nexists C \in \mathcal{CS} \cup \{\emptyset\} : C \ \cup \{a_i\} \succ_i C_i(\mathcal{CS})$ $\wedge \forall a_j \in C, C \ \cup \{a_i\} \succeq_j C$
Stabilité individuelle contractuelle	$\forall a_i \in N, \nexists C \in \mathcal{CS} \cup \{\emptyset\} : C \ \cup \{a_i\} \succ_i C_i(\mathcal{CS})$ $\wedge \forall a_j \in C, C \ \cup \{a_i\} \succeq_j C$ $\wedge \forall a_k \in C_i(\mathcal{CS}), a_k \neq a_i, C_i(\mathcal{CS}) \setminus \{a_i\} \succeq_k C_i(\mathcal{CS})$
Stabilité du cœur	$\forall a_i \in N, \nexists C \in G_i : C \ \succ_i C_i(\mathcal{CS})$ $\wedge \forall a_j \in C, C \succeq_i C_j(\mathcal{CS})$
Stabilité Pareto-optimale	$\nexists \mathcal{CS}_2 : \forall a_i \in N, C_i(\mathcal{CS}_2) \succeq_i C_i(\mathcal{CS})$ $\wedge \ \exists a_j \in N, C_j(\mathcal{CS}_2) \succ_j C_j(\mathcal{CS})$

TABLE 1.1 – Définition de quelques concepts de solutions de jeux de coalitions hédoniques

Ces concepts de solutions caractérisent donc différentes notions de stabilité. Par exemple, le concept de stabilité au sens de Nash contient l'ensemble des solutions $\mathcal{CS}$ pour lesquelles il n'existe aucune coalition qu'un agent préfèrerait rejoindre à la place de sa coalition actuelle $C_i(\mathcal{CS})$. Le concept de solutions modélisant la stabilité au sens du cœur est quant à lui l'ensemble des solutions $\mathcal{CS}$ pour lesquelles aucun groupe d'agent ne souhaite former une autre coalition qu'ils préfèreraient tous à leurs coalitions actuelles. Un dernier exemple avec le concept de solutions caractérisant la stabilité Pareto-optimale : celui-ci regroupe l'ensemble des solutions $\mathcal{CS}$ telles qu'aucun agent a_i ne souhaite quitter sa coalition actuelle $C_i(\mathcal{CS})$ pour une autre coalition qu'il préfère sans dégrader la solution pour au moins un autre agent a_j. Résoudre un jeu de coalitions à utilité non-transférable est un problème NP-complet [Elkind et Wooldridge, 2009]. Certains algorithmes s'appuient sur la maximisation du bien-être égalitaire pour la résolution du jeu [Morge et Nongaillard, 2017].

1.2.2 Jeux à utilité transférable

Une autre vision des jeux de coalitions est celle à utilité transférable [Morgenstern et Von Neumann, 1953, Chalkiadakis *et al.*, 2011]. Lorsque les agents coopèrent, leur coalition reçoit une utilité qui doit désormais être distribuée entre les agents. L'utilité qu'une coalition produit est également définie par une fonction caractéristique, mais à la différence des jeux de coalitions quantitatifs, par une unique valeur réelle au lieu du vecteur de valeurs dans ces derniers.

Définition 1.7 (Jeu de coalitions à utilité transférable)**.** *Un* jeu de coalitions à utilité transférable *est un tuple* $\mathcal{G} = \langle N, v \rangle$ *où* $N = \{a_1, \ldots, a_n\}$ *est un ensemble d'agents, et* $v : 2^N \to \mathbb{R}$ *est la fonction caractéristique qui indique l'utilité* $v(C)$ *de chaque coalition* $C \subseteq N$*.*

1.2.2.1 Distribution des gains

Les agents doivent donc s'accorder sur une distribution des gains des coalitions formées. La somme des gains des agents d'une coalition doit donc être inférieure ou égale à l'utilité de cette coalition. Lorsque la somme des gains des agents d'une coalition est égale à l'utilité de la coalition, alors la coalition respecte une propriété d'*efficacité*. Une fois la distribution choisie, celle-ci est intégrée à un vecteur de gains, également appelé une *imputation*, qui, une fois complet et associé à une structure de coalitions, forme une solution à un jeu à utilité transférable. Une imputation est définie comme suit.

Définition 1.8 (Imputation)**.** *Une* imputation *dans une structure de coalitions* $\mathcal{CS}$ *au sein d'un jeu* $\mathcal{G}$ *est un vecteur de gains tel que* $\vec{x} = \{x_1, \ldots, x_n\}$ *où* x_i *est le gain de l'agent* a_i*, et* $x_i \geq 0$*. Nous notons* $x(C)$ *la somme des gains des agents* a_j *appartenant à la coalition* C *:*

$$x(C) = \sum_{a_j \in C} x_j$$

Une *solution* est donc définie comme suit.

Définition 1.9 (Solution d'un jeu de coalitions)**.** *Une* solution *à un jeu de coalitions* $\mathcal{G}$ *est un tuple* $S_\mathcal{G} = \langle \mathcal{CS}, \vec{x} \rangle$ *où :*

— $\mathcal{CS}$ *est une structure de coalitions de* N*,*

— $\vec{x} = \{x_1, \ldots, x_n\}$ *est une imputation.*

Cependant, l'hypothèse que les agents sont rationnels et égoïstes étant faite, il est assez intuitif de se dire que toutes les solutions ne sont pas acceptables par les agents.

Exemple 10. *Soient $\mathcal{G} = \langle N, v \rangle$ un jeu, avec $N = \{a_1, a_2, a_3\}$ et v telle que :*

$$v = \{(a_1) = 0.83\ ;\ (a_2) = 0.74\ ;\ (a_3) = 0.54\ ;$$
$$(a_1, a_2) = 2.02\ ;\ (a_1, a_3) = 0.81\ ;\ (a_2, a_3) = 1.51\ ;$$
$$(a_1, a_2, a_3) = 3.65\ ;\ (\emptyset) = 0\}$$

Soit une solution $S = \langle \mathcal{CS}, \vec{x} \rangle$ avec $\mathcal{CS} = [(a_1); (a_2, a_3)]$ et $\vec{x} = (0.83, 1.01, 0.50)$. Cette solution S n'est certainement pas acceptable pour a_3 car il gagne moins que dans sa coalition singleton, cette solution ne serait donc pas rationnelle pour lui.

Une solution est donc acceptable dans un jeu à utilité transférable si elle est stable – pour rappel, aucune déviation possible – et respecte les propriétés d'efficacité et de rationalité. Tout comme dans les jeux à utilité non-transférable, des concepts de solutions définissent plusieurs notions de stabilité. D'un point de vue général, résoudre un jeu de coalitions à utilité transférable est un problème NP-complet.

Il existe deux types de concepts de solutions dans ce cadre, à savoir les concepts *ensemblistes* – c'est-à-dire qui définissent un ensemble de solutions acceptables – et les concepts *singleton*, qui définissent une unique imputation, par exemple une valeur de contribution moyenne au sein du jeu.

1.2.2.2 Concepts de solutions singleton

Les deux concepts singletons majeurs sont la *valeur de Shapley* [Kurz, 1988, Shapley, 1953] et l'*indice de Banzhaf* [Banzhaf III, 1964]. Ces concepts caractérisent une valeur unique pour chaque agent, cette valeur devant décrire la contribution marginale moyenne de cet agent au sein du jeu de coalitions. Ces valeurs uniques permettent ensuite de calculer une seule imputation par structure de coalitions (en répartissant l'utilité des coalitions à ses membres, proportionnellement aux valeurs uniques calculées), il devient donc simple de déterminer quelle solution sera choisie. Pour cela, la solution doit respecter certains critères, par exemple un agent ne doit pas gagner moins que l'utilité de sa coalition singleton, ce qui irait à l'encontre de la propriété de rationalité. Ensuite, la solution choisie est celle qui produit le plus d'utilité globale, que nous appelons *bien-être social*. Cependant, calculer cette contribution marginale requiert donc de parcourir l'ensemble de la fonction

caractéristique (et donc les valeurs des coalitions). La différence entre ces deux concepts singleton est que la valeur de Shapley prend en compte plusieurs fois chaque coalition (car considère plusieurs ordres d'arrivées possibles des agents dans les coalitions) tandis que l'indice de Banzhaf ne prend pas en compte qu'une fois chaque coalition (considère donc qu'il n'y a pas d'ordre d'arrivée dans une coalition). Ainsi, l'indice de Banzhaf est moins complexe à calculer mais il perd certaines propriétés intéressantes comme l'efficacité ou l'additivité (la somme des indices de Banzhaf de deux jeux n'est pas égale à l'indice de Banzhaf de la somme de ces deux jeux), sauf dans sa forme normalisée qui permet de garder l'efficacité [van Der Laan et van Den Brink, 1998]. Or, ces propriétés sont intéressantes lorsque par exemple les agents peuvent évoluer dans différents jeux avec différentes fonctions caractéristiques (que nous pouvons voir comme la description de plusieurs contextes avec les mêmes agents, par exemple si les agents échangent leurs postes). Dans un cas où il existe deux jeux différents mais possédant les mêmes agents, la somme des valeurs de Shapley des deux jeux distincts est égale aux valeurs de Shapley de la fusion des deux jeux, cela respecte ainsi la propriété d'additivité. La valeur de Shapley est donc définie comme suit.

Définition 1.10 (Valeur de Shapley)**.** *Soit* Π *l'ensemble des ordres d'arrivée des agents dans les coalitions et* $\pi(a_i)$ *l'ensemble des agents précédent l'agent* a_i *dans un ordre donné. La valeur de Shapley* ϕ_i *est définie par :*

$$\phi_i = \frac{1}{n!} \sum_{\pi(a_i) \in \Pi} v(\pi(a_i) \cup a_i) - v(\pi(a_i))$$

Comme expliqué précédemment, l'indice de Banzhaf ne prend pas en compte l'ordre d'arrivée des agents, sa définition s'en retrouve donc simplifiée.

Définition 1.11 (Indice de Banzhaf)**.** *L'indice de Banzhaf* ψ_i *d'un agent* a_i *est définie par :*

$$\psi_i = \frac{1}{2^{n-1}} \sum_{C \subseteq N \setminus a_i} v(C \cup a_i) - v(C)$$

Il existe cependant d'autres concepts de solutions singleton, possédant une autre sémantique. Tandis que les valeurs de Shapley et Banzhaf calculent une valeur unique reposant sur un principe d'équité, les concepts de la famille des *valeur de solidarité* [Calvo et Gutiérrez, 2010, Nowak et Radzik, 1994, Xu *et al.*, 2016] proposent de calculer une valeur unique basée sur un principe d'égalité. Alors que les répartitions équitables s'appuient sur les contributions marginales individuelles des agents, les répartitions égalitaires

s'appuient sur les contributions marginales moyennes au sein d'une coalition donnée, ce qui rend l'utilité obtenue par les agents davantage dépendante des synergies entre ces derniers.

Définition 1.12 (Valeur de solidarité)**.** *La contribution marginale moyenne d'une coalition C est donnée par :*

$$\mathcal{A}(C) = \frac{1}{|C|} \sum_{a_k \in C} [v(C) - v(C \setminus a_k)]$$

La valeur de solidarité χ_i *de l'agent a_i est alors :*

$$\chi_i = \sum_{C \ni a_i} \frac{(n - |C|)!(|C| - 1)!}{n!} \mathcal{A}(C)$$

Ci-dessous sont montrées les différences entre ces valeurs sur un exemple succinct.

Exemple 11. *Soit $\mathcal{G}$ le jeu présenté dans l'exemple 10 (page 19), les différentes valeurs singletons ainsi que les imputations pour la grande coalition – celle qui est formée – sont montrées par la table 1.2.*

	Shapley	Banzhaf	Val. Sol.
a_1	1.248	1.130	1.238
a_2	1.555	1.435	1.332
a_3	0.848	0.730	1.080
Imputations	$\{1.25; 1.55; 0.85\}$	$\{1.25; 1.59; 0.81\}$	$\{1.24; 1.33; 1.08\}$

TABLE 1.2 – Exemple de calcul des valeurs singletons et imputations

1.2.2.3 Cœur et généralisation

L'autre type de concepts de solutions est celui des concepts ensemblistes, qui peuvent donc comprendre plusieurs solutions. Un premier concept de ce type est le cœur [Gillies, 1959, Osborne et Rubinstein, 1994, Shapley et Shubik, 1966], qui est un concept de solutions dont nous avons précédemment vu la version à utilité non-transférable (voir table 1.1). Dans le cadre transférable, le principe reste le même : ce concept définit l'ensemble des solutions pour lesquelles aucun groupe d'agents ne souhaite dévier collectivement afin de former une nouvelle coalition dont l'utilité est supérieure à la somme des gains de ces agents dans la solution évaluée.

Définition 1.13 (Cœur)**.** *Une solution* $(\mathcal{CS}, \vec{x})$ *appartient au cœur si, et seulement si :*

$$\forall C \subseteq N, x(C) \geq v(C)$$

Le concept de solution du cœur est fort, c'est-à-dire que les contraintes qu'il impose sont difficiles à satisfaire et il peut donc être vide pour un jeu de coalitions donné. Il existe cependant une variante non-vide de celui-ci : l'ϵ-cœur. Cette variante permet de relaxer le cœur de la valeur ϵ, telle une partie d'utilité sacrifiée par les agents pour assurer la stabilité.

Définition 1.14 (ϵ-cœur)**.** *Une solution* $S_{\mathcal{G}} = \langle \mathcal{CS}, \vec{x} \rangle$ *appartient à l'*ϵ*-cœur si, et seulement si :*

$$\forall C \subseteq N, x(C) \geq v(C) - \epsilon$$

Pour un jeu, l'ϵ-cœur non-vide pour lequel ϵ est minimal (0 si le cœur n'est pas vide) est également appelé *dernier cœur* [Maschler *et al.*, 1979, Mochaourab et Jorswieck, 2014].

1.2.2.4 Noyau

Le noyau [Chalkiadakis *et al.*, 2011, Davis et Maschler, 1965] est un concept de solutions fondé sur l'équilibre des excès entre les agents : pour toute paire d'agents, pour une solution donnée et pour chaque coalitions auxquelles appartient le premier agent et n'appartient pas le deuxième, l'excès est la différence maximale entre l'utilité des coalitions et la somme des gains des agents la composant dans la solution. Il définit l'ensemble des solutions pour lesquelles aucun agent ne peut demander à un autre une part de ses gains sous peine de dévier. Pour cela, l'excès entre chaque agent est calculé comme suit.

Définition 1.15 (Excès)**.** *Étant donné une solution* $S_{\mathcal{G}} = \langle \mathcal{CS}, \vec{x} \rangle$*, l'excès de l'agent* a_i *sur l'agent* a_j *dans le vecteur* $\vec{x}$ *est calculé comme étant :*

$$S_{i,j}(\vec{x}) = \max\{v(C) - x(C) | C \subseteq N, a_i \in C, a_j \notin C \}$$

Le noyau contient donc l'ensemble des solutions qui respectent un équilibre d'excès entre les paires d'agents : soit les deux ont le même excès sur l'autre, soit si un agent à plus d'excès que l'autre, alors le gain de ce dernier doit être le montant de sa coalition singleton.

Définition 1.16 (Noyau)**.** *Une solution* $S_{\mathcal{G}} = \langle \mathcal{CS}, \vec{x} \rangle$ *appartient au noyau si, et seulement si, pour toute paire* (a_i, a_j)*, une des trois conditions suivantes est respectée :*

1. $S_{i,j}(x) = S_{j,i}(x)$,
2. $S_{i,j}(x) > S_{j,i}(x)$ *et* $x_j = v(\{a_j\})$,
3. $S_{i,j}(x) < S_{j,i}(x)$ *et* $x_i = v(\{a_i\})$.

Des travaux proposent que les agents puissent transférer l'utilité en excès entre les coalitions afin de rendre stable les structures de coalitions produisant le plus d'utilité globale [Airiau et Sen, 2010].

1.2.2.5 Nucléole

Le nucléole [Maschler *et al.*, 1979, Schmeidler, 1969] est un concept de solutions fondé sur la notion de déficit, c'est-à-dire, pour une coalition et une solution données, la différence entre l'utilité de la coalition et la somme des gains des agents la composant dans la solution. Pour chaque solution, un vecteur de déficit de toutes les coalitions de la fonction caractéristique est créé. Ce vecteur de déficit a donc pour taille $2^{|N|}$, et est ensuite ordonné du déficit le plus grand au plus petit.

Définition 1.17 (Vecteur de déficit)**.** *Pour une solution $S_{\mathcal{G}} = \langle \mathcal{CS}, \vec{x} \rangle$, le déficit d'une coalition C est défini comme étant $d(\vec{x}, C) = v(C) - x(C)$. Le vecteur de déficit $\vec{d}$, ordonné lexicographiquement du déficit le plus grand au plus petit (noté $\succeq_{lex}$ sur l'ensemble ci-après), pour cette solution est donc défini comme suit :*

$$\vec{d}(\vec{x}) = \{d(\vec{x}, C) | \forall C \in 2^N\}^{\succeq_{lex}}$$

Ensuite, ces déficits sont comparés lexicographiquement. Le nucléole est l'ensemble des solutions ayant le vecteur de déficit ordonné le plus petit lexicographiquement [Schmeidler, 1969].

Définition 1.18 (Nucléole)**.** *Soit $\mathcal{I}(\mathcal{G})$ l'ensemble des imputations (vecteurs de gains) possibles pour le jeu de coalitions $\mathcal{G}$, le nucléole est défini comme suit :*

$$\mathcal{N}(\mathcal{G}) = \{\vec{x} \in \mathcal{I}(\mathcal{G}) | d(\vec{x}) \leq_{lex} d(\vec{y}), \forall \vec{y} \in \mathcal{I}(\mathcal{G})\}$$

Il existe également des inclusions entre les concepts de solutions ensemblistes, mises en lumière par la figure 1.1. Cela signifie que l'ensemble des solutions d'un certain concept peut être contenu dans un autre concept. Un concept strictement inclus dans un autre est donc plus strict.

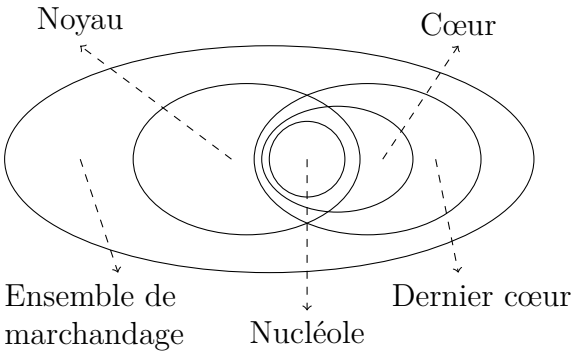


FIGURE 1.1 – Inclusions entre concepts ensemblistes

1.3 Utilité transférable : métriques et jeux spécifiques

Ce qui a été présenté précédemment est ce que nous pouvons appeler la formation de coalitions classique, que ce soit avec ou sans utilité transférable. Les agents utilitaristes ainsi que le partage des utilités des coalitions étant nos centres d'intérêt, intéressons-nous désormais à des aspects plus spécifiques du cadre à utilité transférable, notamment l'évaluation de la qualité d'une solution, ainsi que les extensions possibles aux jeux de coalitions qui permettent de modéliser davantage de contextes réels, comme l'ajout de capacités quantitatives ou qualitatives (par exemple le permis d'utiliser une grue sur un port maritime) ou encore les dépendances entre les coalitions, c'est-à-dire l'influence sur une coalition par les autres au sein d'une structure (par exemple un groupe de dockers sera plus ou moins efficace selon les autres groupes formés).

1.3.1 Évaluer la qualité d'une solution

Lorsqu'une solution est choisie et que les coalitions sont formées, nous pouvons nous interroger sur la possibilité de quantifier son intérêt ou sa pertinence. Cependant, cela dépend des propriétés des agents. Des agents égoïstes auront un intérêt individuel à former des coalitions, tandis que celles-ci seront un moyen d'améliorer l'utilité globale pour des agents utilitaristes et altruistes [1].

1. Comme mentionné précédemment, l'altruisme s'oppose à la notion d'égoïsme.

1.3.1.1 Intérêt global : le bien-être social

Lorsque les agents ne pas égoïstes, la formation de coalitions peut se résumer à la maximisation du bien-être social, la coopération servant donc à augmenter la somme des gains des agents, ceux-ci n'ayant aucune exigence sur leurs gains personnels. Nous pouvons mesurer le bien-être social simplement en faisant la somme des utilités des coalitions d'une structure de coalitions.

Définition 1.19 (Bien-être social)**.** *Le bien-être social $\mathcal{W}$ d'une structure de coalitions $\mathcal{CS}$ est défini comme suit :*

$$\mathcal{W}(\mathcal{CS}) = \sum_{C \in \mathcal{CS}} v(C)$$

Si ce bien-être social pour une structure donnée est maximal dans le jeu, nous pouvons alors parler de *structure de coalitions optimale.*

Définition 1.20 (Structure de coalitions optimale)**.** *Soient un jeu de coalitions $\mathcal{G} = \langle N, v \rangle$ et $\Pi = \{\mathcal{CS}_1, \ldots, \mathcal{CS}_n\}$ l'ensemble des structures de coalitions de l'ensemble d'agents N. Une* structure de coalitions optimale *$\mathcal{CS}^*$ est définie tel que :*

$$\mathcal{CS}^* = \underset{\mathcal{CS}_k \in \Pi}{\operatorname{argmax}} \mathcal{W}(\mathcal{CS}_k)$$

1.3.1.2 Intérêt individuel : la stabilité

En revanche, si les agents sont rationnels, ceux-ci souhaitent gagner au moins autant que ce qu'ils gagneraient dans leurs coalitions singletons. Si les agents sont égoïstes en plus d'être rationnels, ceux-ci voudront obtenir le plus de gains possible. Cela nous ramène donc à la notion de stabilité que les concepts de solutions caractérisent. Par exemple, le concept du cœur respecte la rationnalité des agents car la coalition singleton d'un agent rendra instable toute solution où sa part d'utilité est inférieure à celle produite par sa coalition singleton, tandis que l'égoïsme est traduit par le fait qu'un groupe d'agents déviera pour former une autre coalition où ils gagneraient collectivement plus que dans la solution proposée, comme dans l'exemple 10 (page 19).

Si une solution est stable, l'intérêt individuel est respecté. Cependant dans certains contextes, comme par exemple lorsque le système est décentralisé, les agents peuvent être amenés à former une structure non stable alors qu'il existe une solution stable, et cela en raison de la localité des calculs. Il devient donc intéressant de pouvoir quantifier la distance à la stabilité d'une solution par rapport à une solution stable. À notre connaissance, il

n'existe pas de définition précise de telles mesures dans la littérature, mais la nature même des concepts de solutions permet de caractériser cette distance. Par exemple pour le concept du nucléole, les solutions stables partagent toutes un même vecteur de déficit unique. Il est alors naturel de parler de distance lexicographique entre le vecteur de déficit des solutions stables et le vecteur de déficit d'une solution non stable. Concernant le cœur, la distance peut être simplement donnée par une différence entre l'epsilon du dernier cœur et l'epsilon de l'ϵ-cœur auquel appartient la solution évaluée.

Exemple 12. *Prenons par exemple deux solutions S_1 et S_2 pour le jeu $\mathcal{G}$ de l'exemple 10 (page 19) et le concept du cœur, avec lequel $S_1 = ([(a_1, a_2, a_3)], \{1.00, 2.11, 0.54\})$ est stable, et $S_2 = ([(a_1, a_2, a_3)], \{1.04, 2.11, 0.50\})$ non. La solution S_2 fait partie d'un ϵ-cœur avec $\epsilon = 0.04$. Nous pouvons donc dire que S_2 est non-stable avec une distance de* 0.04.

1.3.1.3 Prix de la stabilité

Cependant même avec des agents égoïstes, le bien-être social peut rester important, bien que relégué au second plan derrière l'individualisme des agents. Nous pouvons souhaiter analyser la perte qu'engendre l'égoïsme des agents, du point de vue de l'intérêt global, au profit de la stabilité. Pour cela, le *prix de la stabilité*, défini par Anshelevich *et al.* [Anshelevich *et al.*, 2008], peut être utilisé. Cette mesure compare simplement le bien-être social d'une solution stable au bien-être social maximal du jeu sans prendre en compte la stabilité.

Définition 1.21 (Prix de la stabilité)**.** *Étant donné un jeu $\mathcal{G}$, soient $\mathcal{CS}^*_{\mathcal{G}}$ une structure de coalitions optimale (voir Définition 1.20) du jeu $\mathcal{G}$, et $S_{\mathcal{G}} = \langle \mathcal{CS}, \vec{x} \rangle$ une solution quelconque au jeu $\mathcal{G}$. Le* prix de la stabilité *pour cette solution, noté $PS(S_{\mathcal{G}})$, est défini par :*

$$PS(S_{\mathcal{G}}) = \frac{\sum_{C \in \mathcal{CS}} v(C)}{\sum_{C' \in \mathcal{CS}^*_{\mathcal{G}}} v(C')}$$

1.3.2 Jeux de coalitions et fonctions caractéristiques spécifiques

La formation de coalitions présentée en section 1.2.2 est la forme la plus classique possible, sans restrictions sur la fonction caractéristique. La littérature étudie des variantes des jeux de coalitions, où les différences par rapport à la formation de coalitions classique se situent soit sur la structure de la fonction caractéristique, soit sur des modifications des propriétés des jeux de coalitions, soit sur des ajouts d'éléments à ces jeux. Ci-dessous,

nous présentons certains de ces modèles proposés dans la littérature [Chalkiadakis *et al.*, 2011].

1.3.2.1 Différentes caractérisations de la fonction caractéristique

En premier lieu, nous pouvons parler de la structure des fonctions caractéristiques. Dès la proposition des premiers modèles de jeux de coalitions, certaines structures spécifiques ont également été étudiées. Celles-ci possèdent parfois des propriétés simplifiant la résolution des jeux associés.

Une première structure de fonction caractéristique peut être trouvée dans les *jeux monotones*. La propriété de *monotonicité* de la fonction caractéristique représente le fait que lorsqu'un agent rejoint une coalition, l'utilité de celle-ci ne peut pas baisser.

Définition 1.22 (Fonction caractéristique monotone)**.** *Une fonction caractéristique est dite* monotone *si pour toute paire de coalitions* $C_1, C_2 \subseteq N$ *telles que* $C_1 \subseteq C_2$*, alors* $v(C_1) \leq v(C_2)$*.*

Une autre propriété applicable à la fonction caractéristique est celle de la *superadditivité*, dans les *jeux superadditifs*. Cette propriété est plus forte que la monotonicité car elle implique celle-ci. La superadditivité définit que la coalition résultant de l'union de deux coalitions ne produira jamais moins d'utilité que la somme des utilités des deux coalitions avant leur fusion.

Définition 1.23 (Fonction caractéristique superadditive)**.** *Une fonction caractéristique est dite* superadditive *si pour toute paire de coalitions* $C_1, C_2 \subseteq N$*, alors* $v(C_1 \cup C_2) \geq v(C_1) + v(C_2)$*.*

Un dernier exemple est celui des *jeux simples*. Ceux-ci sont des jeux monotones où les coalitions ne peuvent prendre que deux valeurs : 0 ou 1. Une coalition ayant une valeur de 1 est dite *gagnante* tandis qu'une coalition avec une valeur de 0 est dite *perdante*. Ces jeux peuvent représenter simplement des cadres où des tâches sont à réaliser, et seules les coalitions gagnantes représentent celles à même de réaliser une des tâches.

Définition 1.24 (Fonction caractéristique simple)**.** *Une fonction caractéristique est dite* simple *si pour toute coalition* $C \subseteq N$*, alors* $v(C) \in \{0, 1\}$*.*

D'autres types de jeux pouvant modéliser des tâches explicites ont été proposés par la suite dans la littérature.

1.3.2.2 Jeux à capacités

Dans les jeux de coalitions classiques, les tâches à réaliser sont abstraites : elles ne sont pas représentées et n'existent qu'à travers un ensemble d'agents et l'utilité qu'ils tirent des coalitions. Or, dans une situation plus complexe, les agents peuvent avoir des capacités spécifiques que l'on peut exprimer formellement, et ils peuvent devoir réaliser des tâches précises qui requièrent une collection de compétences spécifiques. Dans le même ordre d'idée, les agents peuvent être aussi soumis à des contraintes de ressources non-renouvelables : exécuter certaines tâches ou utiliser certaines capacités peut être coûteux pour les agents.

Afin que les jeux de coalitions puissent modéliser des systèmes comprenant des tâches requérant des capacités particulières, qu'elles soient décrites de façon qualitative ou quantitative, il a été proposé des jeux à compétences [Bachrach et Rosenschein, 2008, Chalkiadakis et Boutilier, 2004, Chalkiadakis et Boutilier, 2008, Gaston et desJardins, 2005, Ohta *et al.*, 2006, Wooldridge et Dunne, 2006] et des jeux à ressources [Aubin, 1981, Borkotokey et Neog, 2014, Kraus *et al.*, 2003, Shehory et Kraus, 1998, Wooldridge et Dunne, 2006, Xu *et al.*, 2017] que nous regroupons sous l'appellation *jeux à capacités*. Les tâches devenant explicites, il est ajouté au jeu un ensemble de tâches, que les coalitions doivent réaliser pour en tirer un gain. La fonction caractéristique est alors non plus définie par les coalitions mais par les tâches.

Définition 1.25 (Jeu à capacités)**.** *Un* jeu de coalitions à capacités *est un tuple* $\mathcal{G} = \langle N, \mathcal{T}, v \rangle$ *où :*

- $N = \{a_1, \ldots, a_n\}$ *est un ensemble d'agents,*
- $\mathcal{T} = \{t_1, \ldots, t_m\}$ *est un ensemble de tâches,*
- $v : 2^{\mathcal{T}} \to \mathbb{R}$ *est une fonction caractéristique qui associe à chaque ensemble de tâches une valeur réelle qui est l'utilité obtenue par une coalition lorsque qu'elle les réalise.*

Les agents sont quant à eux dotés de capacités. Une capacité qualitative doit être vue comme une compétence pouvant être requise pour la réalisation d'une tâche [Bachrach et Rosenschein, 2008]. Une capacité quantitative peut être vue comme une fiabilité de l'agent, une ressource qu'il possède et peut mettre à profit pour la réalisation d'une tâche, ou bien comme une capacité de stockage [Shehory et Kraus, 1998].

Dans le cadre qualitatif, il a été proposé les *jeux à compétences*. Dans ceux-ci, un ensemble de compétences $S = \{s_1, \ldots, s_k\}$ est ajouté au jeu, et les agents possèdent chacun

un sous-ensemble $S(a_i) \subset S$ de compétences. À chaque tâche est également associé un sous-ensemble $S(t_j) \subset S$, décrivant l'ensemble des compétences requises pour réaliser cette tâche. Afin de maximiser leur utilité, les agents doivent former des coalitions réunissant les compétences nécessaires à la réalisation des tâches.

Définition 1.26 (Pouvoir qualitatif)**.** *Le* pouvoir qualitatif *d'une coalition C est défini par $S(C) = \bigcup_{a_i \in C} S(a_i)$. Une coalition C peut réaliser une tâche t_j si, et seulement si, les compétences nécessaires à la tâche sont incluses dans son pouvoir, i.e. $S(t_j) \subseteq S(C)$.*

Des sous-classes de ce modèle ont été considérées, comme les *task count skill games*, où la valeur d'une coalition est égale au nombre de tâches qu'elle peut accomplir, et les *weighted task skill games* où elle est égale à la somme de poids associés à chaque tâche [Bachrach et Rosenschein, 2008]. Résoudre un jeu à compétences est un problème NP-difficile [Aziz et De Keijzer, 2011] (plus complexe qu'un jeu de coalitions classique, pour rappel NP-complet) sauf dans le cas favorable où les tâches requièrent un nombre constant de compétences, où cela devient polynomial [Bachrach *et al.*, 2010].

Dans le cadre quantitatif, il a été proposé des *jeux à ressources* comme illustré par le modèle de Shehory et Kraus [Shehory et Kraus, 1998]. Ces jeux considèrent un ensemble de capacités quantitatives $\mathcal{B} = \{b_1, \ldots, b_r\}$. Les agents ainsi que les tâches sont associés à des vecteurs qui décrivent les ressources disponibles et les ressources requises : $\forall a_i \in N, \vec{b_i} = \langle b_1^i, \ldots, b_r^i \rangle$ et $\forall t_j \in \mathcal{T}, \vec{b_{t_j}} = \langle b_1^{t_j}, \ldots, b_r^{t_j} \rangle$.

Définition 1.27 (Pouvoir quantitatif)**.** *Le* pouvoir quantitatif *d'une coalition C est donné par un vecteur $\vec{b_C} = \sum_{a_i \in C} \vec{b_i}$. Une coalition peut réaliser une tâche t_j si, et seulement si, les ressources requises sont inférieures aux ressources disponibles : $\forall k \in [1, r], b_k^{t_j} \leq b_k^C$.*

Il existe une sous-classe des jeux à ressources appelés *jeux flous* [Aubin, 1981, Borkotokey et Neog, 2014, Xu *et al.*, 2017]. Dans ces jeux, il n'y a ni tâches, ni ressources explicites mais, de manière plus abstraite, les agents peuvent choisir leur niveau d'implication dans la coalition (équivalent aux ressources disponibles). La fonction caractéristique est alors dépendante du niveau d'implication de chacun des agents.

Enfin, les jeux de coalitions classiques ne prennent pas en compte les restrictions que les agents peuvent avoir en termes de communication ou d'accointance : tous les agents peuvent former des coalitions avec tous les autres. Or, dans certains domaines, comme les réseaux pair-à-pair ou les chaînes logistiques [Apt et Witzel, 2009, Shehory et Kraus, 1998], les agents sont soumis à de telles contraintes et où ils possèdent généralement

des capacités quantitatives, comme par exemple une capacité de transport limitée. Pour modéliser de tels problèmes, il a été proposé les *network flow games* (NFG) [Deng *et al.*, 2009, Granot et Granot, 1992, Kalai et Zemel, 1982, Kern et Paulusma, 2003]. Ces modèles représentent la fonction caractéristique par un graphe de flot où les agents sont des arcs et leur capacité est naturellement le flot qui les traversent. La valeur d'une coalition C est alors égale au flot maximum passant uniquement par les agents de la coalition. De manière intéressante, calculer le cœur ou le nucléole d'un NFG simple où tous les agents ont une capacité de 1 est un problème polynomial, mais cela devient NP-difficile dans les autres cas [Deng *et al.*, 2009, Kern et Paulusma, 2003].

1.3.2.3 Jeux à coalitions recouvrantes

Dans un jeu à compétences, l'indépendance entre les coalitions peut être une limite, par exemple si un agent est le seul à posséder une compétence requise pour plusieurs tâches différentes, il ne pourra participer qu'à une seule coalition. De même, si le pouvoir d'une coalition d'un jeu à ressources surpasse celui requis pour réaliser ses tâches, alors les ressources supplémentaires sont perdues. Ces deux exemples posent une question commune : est-il possible que les agents puissent participer à plusieurs coalitions ? Pour pallier cette limite, la littérature propose les *jeux à coalitions recouvrantes* [Chalkiadakis *et al.*, 2008, Shehory et Kraus, 1996] qui permettent aux agents de participer à plusieurs coalitions simultanément.

Définition 1.28 (Jeu à coalitions recouvrantes)**.** *Un* jeu à coalitions recouvrantes *est un tuple* $\mathcal{G} = \langle N, v \rangle$ *où :*

- $N = \{a_1, \ldots, a_n\}$ *est un ensemble d'agents,*
- $v : [0,1]^n \to \mathbb{R}$ *est une fonction caractéristique qui associe à chaque vecteur de* n *réels dans* $[0,1]$ *un réel, appelé utilité de la coalition et noté* $v(C)$ *où* $C \subseteq N$*.*

Ces vecteurs de taille n représentent des coalitions, appelées *coalitions recouvrantes*, dont les agents peuvent faire plus ou moins partie. Chaque composante de ces vecteurs décrit donc le niveau d'implication d'un agent donné comme la fraction de ses ressources qu'il attribue à cette coalition. Un agent n'attribuant aucune ressource à une coalition n'en fait donc pas partie. Le nombre d'agents participant à une coalition est appelé le *support* de cette dernière.

Le cadre classique ne permettant pas de décrire une solution dans un jeu de coalitions recouvrantes, les notions d'imputation et de solution, et par extension les concepts de

solutions, doivent être redéfinis.

Une imputation pour une structure de coalitions $\mathcal{CS}$ (de taille k) est désormais définie par un tuple de k vecteurs de taille n, chacun décrivant la répartition de la valeur de la coalition parmi son support. Les agents qui n'y participent pas ont une valeur nulle. Le gain x_i d'un agent a_i dans une imputation $\vec{x}$ est alors la somme de ses gains dans chaque vecteur du tuple. L'ensemble de toutes les imputations pour la structure de coalitions $\mathcal{CS}$ est notée $I(\mathcal{CS})$. Une solution à un jeu à coalitions recouvrantes est donc simplement un tuple consistué d'une structure de coalitions $\mathcal{CS}$ et d'une imputation $\vec{x} \in I(\mathcal{CS})$.

Ainsi, le cœur devient comme suit :

Définition 1.29 (Cœur recouvrant)**.** *Une solution $\langle \mathcal{CS}, \vec{x} \rangle$ est dans le* cœur recouvrant *si pour tout ensemble d'agents $C \subseteq N$, pour toute structure de coalitions $\mathcal{CS}_C$ pour C – c'est-à-dire l'ensemble des structures de coalitions contenant C – et pour toute imputation $\vec{y} \in I(\mathcal{CS}_C)$, nous avons $y_i(\mathcal{CS}_C, \vec{y}) \leq x_i(\mathcal{CS}, \vec{x})$ pour tout agent $a_i \in C$.*

La complexité des jeux recouvrants est dépendante du montant initial de ressources des agents, de la taille maximale des coalitions et de l'ordre des interactions entre les agents. Le problème de décision associé reste NP-difficile [Zick *et al.*, 2012].

1.3.2.4 Jeux à externalités

Dans certaines situations, les coalitions peuvent influencer les autres. Par exemple, elle peuvent entrer en conflit avec d'autres coalitions lorsqu'elles doivent utiliser des outils ou réaliser des tâches en exclusion mutuelle. Ce type de situation est modélisé par les *jeux à externalités*, aussi appelés *partition function games* (PFG) [Thrall et Lucas, 1963]. Les PFGs définissent une notion de *coalitions intégrées* qui caractérise les coalitions dont l'utilité dépend de la structure de coalitions dans laquelle elles apparaissent.

Définition 1.30 (Coalition intégrée)**.** *Une* coalition intégrée *est un tuple $(C, \mathcal{CS})$ où $C \in \mathcal{CS}$ et $\mathcal{CS}$ une structure de coalitions. L'ensemble des coalitions intégrées de N est noté E_N. La fonction caractéristique est redéfinie comme suit : $v : E_N \to \mathbb{R}$. Ainsi, $v(C, \mathcal{CS})$ est la valeur de la coalition intégrée C dans $\mathcal{CS}$.*

Si les concepts de solutions ensemblistes classiques peuvent être facilement redéfinis dans le cadre des PFGs [Chalkiadakis *et al.*, 2011, Thrall et Lucas, 1963], cela n'est pas aussi aisé pour les concepts de solutions singleton car les axiomes garantissant l'unicité de ces concepts dans le cadre classique ne le garantissent plus dans le cadre des PFGs

[Michalak *et al.*, 2010]. De plus, calculer des solutions stables devient plus difficile et peu de travaux ont proposé des solutions algorithmiques à ce problème [Michalak *et al.*, 2009, Rahwan *et al.*, 2012, Sklab *et al.*, 2020, Ueda *et al.*, 2012] car la taille de la fonction caractéristique dépend maintenant du nombre de structures de coalitions, et non plus du nombre de coalitions.

1.3.3 Méthodes de résolution

Les approches courantes de résolution de jeux de coalitions, qu'ils soient classiques ou qu'ils en soient des variantes, se fondent généralement sur des processus centralisés, exacts ou approchés avec une diversité de méthodes : programmation dynamique, programmation linéaire ou méthodes méta-heuristiques [Rahwan *et al.*, 2015]. Cependant, des travaux sont consacrés depuis quelques temps à la résolution décentralisée des jeux de coalitions, notamment grâce à des algorithmes fondés sur des négociations à partir de préférences calculées localement ou sur des graphes dynamiques.

1.3.3.1 Algorithmes centralisés

Une première approche regroupe les algorithmes de génération de structures de coalitions (CSG), notamment fondés sur celui proposé par Sandholm *et al.* [Sandholm *et al.*, 1999], qui permettent la recherche de la structure de coalitions maximisant le bien-être social. Il n'y a donc aucune notion de stabilité, mais uniquement d'optimalité au sens du bien-être social. Le principe de ces algorithmes repose sur une méthode gloutonne, à savoir l'exploration du treillis des structures de coalitions, en partant des coalitions singleton ou de la grande coalition selon les algorithmes. Ces algorithmes sont développés dans une méthode de programmation dynamique, à commencer par l'algorithme proposé par Yun Yeh [Yun Yeh, 1986], nommé **DP**. Il consiste à débuter l'exploration par les coalitions singleton, en les fusionnant et évaluant le bien-être social suite à cette fusion. Si cela est bénéfique, ce bien-être social est enregistré, puis d'autres fusions sont effectuées avec d'autres éléments, et ainsi de suite. Si ce n'est pas le cas, l'algorithme revient en arrière pour essayer une autre fusion. Au contraire, l'algorithme **IDP** proposé par Rahwan et Jennings [Rahwan et Jennings, 2008], qui est une amélioration de celui de Yun Yeh, consiste à débuter l'exploration par la *grande coalition* (qui contient tous les agents) puis la décomposer et évaluer si cette décomposition est bénéfique pour le bien-être social. Si c'est le cas, le bien-être social est enregisté puis la décomposition continue. Si la décomposition

n'était pas bénéfique, l'algorithme revient en arrière et essaye une autre décomposition. Une autre différence par rapport à l'algorithme précédent est que les auteurs s'appuient également sur les travaux de génération de structure de coalitions *anytime* qui permettent d'obtenir une solution à n'importe quel moment de l'algorithme, sans garantie de l'optimalité [Sandholm *et al.*, 1999]. Nous pouvons donc résumer les étapes de ces protocoles comme suit.

1. Choix de départ (coalitions singleton pour **DP**, grande coalition pour **IDP**) ;
2. Fusion (**DP**) ou décomposition (**IDP**) ;
3. Évaluation : **si bénéfique**, étape 4, **sinon** annulation et retour à l'étape 2 ;
4. Mise en cache du bien-être social et de la solution courante, puis étape 2 ;
5. Arrêt après parcours complet (**DP**) ou *anytime* (**IDP**).

Cependant, ce type de résolution peut devenir difficile en raison de la taille exponentielle du treillis des structures de coalitions. Ce nombre de structures de coalitions peut être calculé grâce au nombre de Bell [Bell, 1938] (qui décrit le nombre de partitions d'un ensemble).

Lorsque la question de la résolution d'un jeu de coalitions avec un concept de solutions précis se pose, l'approche est différente. Celle-ci peut par exemple être la programmation linéaire [Schrijver, 2003]. La programmation linéaire permet de caractériser de façon simple un concept de solutions. Un exemple est celui du programme linéaire caractérisant le cœur d'un jeu de coalitions dont la fonction caractéristique est superadditive [Chalkiadakis *et al.*, 2011].

$$\begin{aligned} x_i &\geq 0 \ \forall a_i \in N \\ \sum_{a_i \ \in N} x_i &= v(N) \\ \sum_{a_i \in C} x_i &\geq v(C) \ \forall C \subseteq N \end{aligned} \tag{1.1}$$

Dans un tel cadre, le programme linéaire a une complexité polynomiale, cependant, cette complexité augmente lorsque la fonction caractéristique n'est pas superadditive, et également si de nouvelles variables sont introduites, comme des capacités pour les agents [Schrijver, 2003].

1.3.3.2 Algorithmes distribués

Enfin, une dernière méthode de génération de structures de coalitions est celle utilisant une méta-heuristique [Mauro *et al.*, 2010]. Dans celle-ci, la recherche est distribuée et non centralisée, mais l'algorithme s'inspire fortement de ce qui a été fait en programmation dynamique centralisée et *anytime* [Rahwan *et al.*, 2009, Sandholm *et al.*, 1999]. Il est donc assumé que l'algorithme produit une solution approximée mais garantie de bonne qualité néanmoins. L'algorithme procède en deux étapes : une construction gloutonne suivie d'une recherche locale perturbative, c'est-à-dire des modifications des composants locaux de la solution évaluée (dans ce cas précis, les agents des coalitions) [Hoos et Stützle, 2004]. Plus précisément, l'algorithme commence avec une solution vide, et ajoute séquentiellement des composants – c'est-à-dire des coalitions – selon une fonction de sélection heuristique. S'ensuit l'étape de recherche locale perturbative pour améliorer localement la solution candidate. L'aspect stochastique de l'algorithme améliore l'exploration du treillis comparé aux algorithmes uniquement gloutons.

1.3.3.3 Algorithmes décentralisés

Nous nous intéresserons ici aux algorithmes décentralisés, plus adaptés à un cadre multi-agents. Ces algorithmes sont souvent proposés pour un contexte applicatif spécifique, se restreignant en particulier aux jeux à ressources [Gaston et desJardins, 2005, Glinton *et al.*, 2008, Mihailescu *et al.*, 2011]. Toutefois, ils partagent des bases communes et deux catégories se distinguent : les algorithmes fondés sur des négociations à partir de préférences calculées localement [Bremer et Lehnhoff, 2017, Diago *et al.*, 2016, Khalouzadeh *et al.*, 2010, Shehory et Kraus, 1995, Shehory et Kraus, 1996, Shehory et Kraus, 1998, Sims *et al.*, 2003], et les algorithmes fondés sur des graphes dynamiques [Bistaffa *et al.*, 2017, Gaston et desJardins, 2005, Glinton *et al.*, 2008, Mihailescu *et al.*, 2011, Voice *et al.*, 2012].

Dans les approches fondées sur la négociation, une première étape consiste à extraire les préférences des agents sur les coalitions possibles, en calculant leur utilité espérée. Par exemple, Shehory et Kraus [Shehory et Kraus, 1995, Shehory et Kraus, 1996, Shehory et Kraus, 1998] établissent ces préférences via un protocole de communication itéré : les agents constituent une liste restreinte de coalitions – d'une taille maximale fixée – qu'ils désirent former et contactent les agents en faisant partie afin de s'enquérir de leurs capacités, ce qui leur permet d'estimer leur valeur. Ci-dessous est décrit le fonctionnement

général de l'algorithme.

1. Chaque agent construit une liste restreinte de coalitions dont il peut faire partie et avec une taille maximale fixée ;
2. Chaque agent contacte les autres agents impliqués dans ses coalitions souhaitées ;
3. En estimant la valeur des autres agents, chaque agent calcule localement une valeur nommée *poids* ;
4. La coalition ayant le poids minimal est formée ;
5. Le processus est répété jusqu'à ce que chaque agent fasse partie d'une coalition.

D'autres approches se passent de communication et s'appuient plutôt sur des heuristiques, soit en donnant une préférence aux agents les plus proches dans des modèles spatiaux [Khalouzadeh *et al.*, 2010, Sims *et al.*, 2003], ou en minimisant les erreurs d'attribution de ressources [Bremer et Lehnhoff, 2017]. Enfin, une fois ces estimations réalisées, la seconde étape consiste à négocier : des coalitions sont proposées (par les agents eux-mêmes ou un commissaire-priseur) et, si tous les membres l'acceptent, elles sont formées. Une dernière approche fondée sur les négociations utilise des préférences liées à des critères : l'utilité des agents est décrite par un ensemble de critères décrivant leurs objectifs. Les agents proposent ensuite des solutions satisfaisant leurs critères, puis des négociations basées sur l'argumentation formelle (principe d'attaque/défense permettant d'argumenter sur des propositions) sont menées pour décider quelle sera la solution finalement retenue [Diago *et al.*, 2016].

Exemple 13. *Dans l'exemple du port maritime, un critère pourrait être vu comme la préférence d'un douanier d'être avec un autre douanier spécifique. Un tier parti pourrait argumenter qu'ils seraient plus efficaces séparés en attaquant la préférence du premier.*

Dans les approches fondées sur des graphes dynamiques, des contraintes en termes de communication sont ajoutées au modèle : les agents ne peuvent pas communiquer avec n'importe quels autres. Pour cela, nous trouvons des représentations sous forme de *graphes contraints* [Bistaffa *et al.*, 2017, Voice *et al.*, 2012] ou de *réseaux d'agents organisés* [Gaston et desJardins, 2005, Glinton *et al.*, 2008, Mihailescu *et al.*, 2011]. Il s'agit essentiellement d'une distinction cosmétique car les deux représentations possèdent un fonctionnement similaire. La structure de coalitions est représentée par un graphe dont les agents en sont les sommets, et les arêtes entre ces sommets représentent une interaction. Une coalition est un sous-graphe connexe du graphe général. Durant chaque étape du protocole, les agents

ne peuvent former des coalitions qu'avec un ensemble de voisins fixés *a priori*, et décident de cela via des heuristiques. Si deux agents décident de rejoindre une même coalition alors une arête est ajoutée entre eux. Si un agent souhaite quitter une coalition, les arêtes entre cet agent et les membres de la coalition sont détruites. Selon le modèle, la formation de coalitions s'arrête soit après un certain temps [Glinton *et al.*, 2008, Mihailescu *et al.*, 2011], soit lorsqu'il n'y a plus de tâches à accomplir [Bistaffa *et al.*, 2017, Gaston et desJardins, 2005, Voice *et al.*, 2012]. Dans ces modèles, la nature des heuristiques influe évidemment grandement sur l'optimalité de la solution. Par exemple, une politique gloutonne sur les performances des agents produit de bons résultats tandis qu'une heuristique encourageant la diversité des capacités est peu satisfaisante [Gaston et desJardins, 2005, Glinton *et al.*, 2008]. Enfin, malgré l'ajout de contraintes de communication réduisant le nombre de structures de coalitions possibles à un instant donné, le problème reste NP-complet [Voice *et al.*, 2012].

Comme souligné ci-dessus, les protocoles décentralisés peinent souvent à trouver et/ou garantir une solution globalement optimale [Bremer et Lehnhoff, 2017, Gaston et desJardins, 2005, Glinton *et al.*, 2008, Shehory et Kraus, 1998]. Cela est notamment au fait qu'ils proposent une résolution approchée du problème, en raison du manque d'information.

1.4 Problématiques

Un des buts des systèmes multi-agents est de modéliser des situations réelles, comme par exemple des chaînes logistiques ou des réseaux de capteurs intelligents. Cependant, des hypothèses faites dans le cadre classique freinent cette mise en application, comme par exemple le déterminisme des utilités, la connaissance *a priori* de ces dernières ou bien l'indépendance des coalitions. Nous avons cependant évoqué beaucoup de modèles précédemment, répondant chacun à une problématique particulière. Nous pouvons faire une taxonomie de ces modèles, selon les hypothèses relâchées dans le but de répondre à certains besoins applicatifs, décrits ci-dessous :

- Agents homogènes : applications impliquant une gestion des ressources distinctes, ou mettant en jeu des agents hétérogènes dans leurs compétences,
- Indépendance : applications comprenant une dimension spatiale dans l'organisation des agents ou la distribution de leurs ressources,
- Déterminisme : applications impliquant de la temporalité ou des événements stochastiques liés aux agents ou à leur environnement.

Nous n'avons pas ici la prétention d'être exhaustifs mais de mettre en lumière la diversité des modèles et leur capacité à répondre aux besoins des systèmes multi-agents dans des contextes applicatifs, tel que pour notre exemple applicatif de port maritime. La taxonomie des modèles est présentée par la figure 1.2. L'origine, marquée d'un point plus épais, représente les modèles classiques tandis que les axes représentent chacun le relâchement d'une hypothèse : agents homogènes (axe bleu), coalitions indépendantes (axe noir), gestion de l'incertitude (axe rouge).

En s'appuyant sur cette taxonomie, nous avons pu identifier quelles hypothèses sont sous-représentées, à savoir celles de non-déterminisme et de dépendance entre les coalitions. Cependant, contrairement à l'hypothèse de non-déterminisme, les modèles intégrant de la dépendance entre coalitions ne sont pas fortement contextualisés, et peuvent s'appliquer rapidement à des contextes applicatifs comme notre port maritime, c'est pour cette raison que nous allons nous concentrer sur la première hypothèse : le non-déterminisme. Un point essentiel évoqué précédemment mais n'apparaissant pas dans la taxonomie concerne les méthodes de résolution. En effet, si divers algorithmes décentralisés sont étudiés dans la littérature, leur application réelle peut être compliquée en raison de leur contextualisation forte au problème pour lequel ils ont été étudiés.

1.4.1 Déterminisme

Une des hypothèses courantes et les plus restrictives dans la formation de coalitions classique est celle du déterminisme de l'utilité créée par les coalitions. En effet, cela implique que si les mêmes agents coopèrent plusieurs fois ensemble, alors le résultat de la coopération sera exactement le même, qu'importe le cadre applicatif. Par exemple, si c'est la réalisation d'une tâche comme un déchargement dans une chaîne logistique, alors cela leur prendra exactement le même temps. Or, dans de nombreux cadres applicatifs, des différences d'efficacité peuvent arriver, provenant parfois de l'environnement ou bien de la forme de l'agent. Il devient donc intéressant d'envisager des modèles où les coalitions ne produiraient pas une utilité déterministe, mais stochastique.

1.4.2 Connaissance *a priori*

Cependant, même si les utilités des coalitions ne sont plus déterministes, les agents ont toujours une connaissance parfaite de la fonction caractéristique. Cela comprend les compétences ou capacités des autres agents, ainsi que les synergies qu'il existe entre eux.

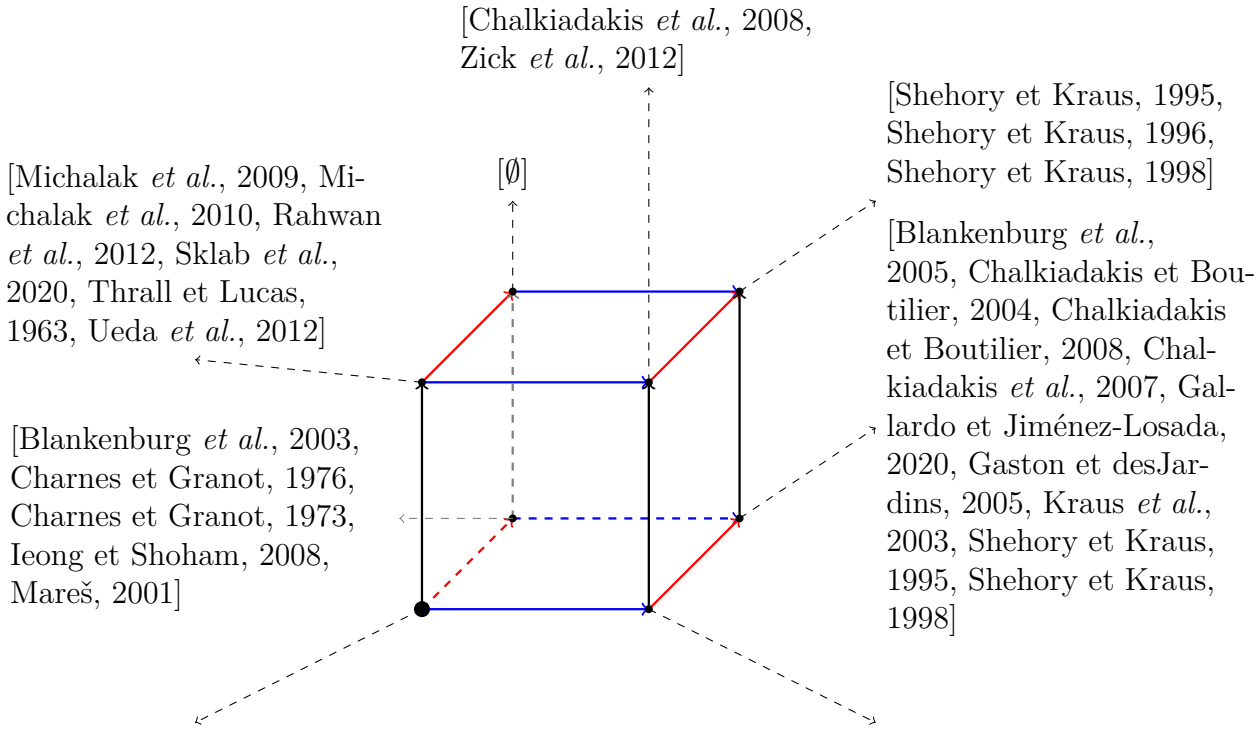


FIGURE 1.2 – Taxonomie des modèles : à partir du cadre classique (point noir épais), l'axe bleu indique la prise en compte des capacités des agents, l'axe noir indique la prise en compte des dépendances entre coalitions et l'axe rouge indique la prise en compte de l'incertitude.

Cela n'est pas non plus compatible avec beaucoup d'applications réelles, notamment qui impliquent des systèmes ouverts. S'affranchir de l'hypothèse de la connaissance *a priori* des agents de la fonction caractéristique implique alors d'introduire une autre sorte d'incertitude dans les jeux de coalitions. Cependant, les agents ne pouvant pas raisonner efficacement en ne connaissant pas la fonction caractéristique réelle, ceux-ci doivent, comme pour l'incertitude produite par la stochasticité, faire une estimation de l'utilité des coalitions. Pour ceci, l'enjeu est l'observation : lorsque des gains sont produits, les agents les observent et peuvent donc estimer les futurs gains. Plus il y a d'observations et donc d'informations, moins l'incertitude sera grande.

1.4.3 Formation de coalitions centralisée

Résoudre un problème de formation de coalitions de manière centralisé n'est pas viable pour certaines applications, telles que les chaînes logistiques ou les réseaux électriques intelligents, où dans ces cas, un nombre important d'agents doit être géré. En effet, rechercher une structure de coalitions optimale, c'est-à-dire qui maximise le bien-être social et qui est stable, dans un cadre centralisé implique d'explorer le treillis des structures, dont la taille croît exponentiellement. Former des coalitions de manière distribuée, voire décentralisée, devient alors intéressant, même si cela implique des redondances ou des coûts de communication et de coordination.

La levée de ces trois hypothèses est donc un enjeu majeur pour permettre de modéliser des problèmes réels grâce à la formation de coalitions, et c'est à ceci que nous allons nous intéresser. Dans un premier temps, dans les chapitres 2, 3 et 4, la levée des hypothèses de déterminisme et de connaissance *a priori* sera notre centre d'intérêt. À cette fin, nous aurons une approche basée sur les bandits manchots, grâce à une analogie que nous pouvons faire entre les bras du bandit manchot et les coalitions, dont les utilités seront donc inconnues et décrites par des distributions de probabilité (tout comme les gains des bras). Cette analogie nous permettra également d'intégrer une notion de l'apprentissage par renforcement, afin de traîter l'absence de connaissance et la stochasticité : l'équilibre exploration-exploitation. Dans un second temps, dans les chapitres 5 et 6 la décentralisation dans un contexte stochastique répété sera notre sujet d'étude, en se basant sur un protocole distribué de concessions monotones. Ce dernier, exempt d'hypothèse sur la structure du système et d'entité centrale, est un parfait candidat pour l'adaptation à la formation de coalitions, et ce même dans un contexte où les hypothèses de déterminisme

et de connaissance *a priori* ont été levées.

Chapitre 2

Formation de coalitions dans l'incertitude et bandits manchots

Sommaire

Une première étape est donc la levée des hypothèses de déterminisme et de connaissance *a priori* qui caractérisent respectivement le fait que les utilités produites par les coalitions sont fixes et que ces mêmes utilités sont connues des agents avant le déroulement du jeu. La levée de telles hypothèses amène évidemment son lot de questions. Comment représenter l'incertitude dans les modèles de formation de coalitions ? Comment les agents peuvent-ils estimer les utilités des coalitions ? Comment peuvent-ils prendre une décision dans un tel contexte ? Dans ce chapitre, nous présentons les travaux de la littérature du domaine de la formation de coalitions dont les questions sont connexes aux nôtres, ainsi qu'un problème de décision séquentielle, les bandits manchots, avec lequel nous pourrons faire une analogie, afin de nous permettre d'utiliser des stratégies de ce problème pour la formation de coalitions dans un cadre stochastique et répété.

2.1 Incertitude dans les jeux de coalitions

Dans la littérature, des travaux ont porté sur la gestion de l'incertitude au sein des jeux de coalitions, et ce de plusieurs manières. En effet, dans un système multi-agents, l'incertitude peut provenir de différentes sources : l'environnement, les capacités des agents, des biais de communications (communications contraintes dans l'espace, coûts ou bruits pour les communications) ou bien juste par un manque d'informations (connaissances locales des agents).

Exemple 14. *Dans un port maritime, certains éléments peuvent venir perturber le bon déroulement des diverses activitées.*

- ***Environnement :*** *Une mauvaise météo peut compliquer la navigation des navires et entraîner du retard ou des difficultés concernant les différentes manoeuvres à effectuer.*
- ***Capacités :*** *Une panne d'un véhicule ou d'une machine peut retarder le chargement ou déchargement de conteneurs. Un autre exemple est si un docker est malade, il travaillera probablement moins efficacement que d'habitude, et mettra plus de temps à effectuer ses tâches.*
- ***Biais de communications :*** *Une mauvaise communication entre un employé et son supérieur peut entraîner la réalisation d'une certaine tâche à la place d'une autre, et ainsi avoir un impact ailleurs.*
- ***Manque d'informations :*** *En l'absence d'informations précises sur le chargement d'un navire, les douaniers peuvent se retrouver en sous-effectif par rapport à la charge de travail et donc prendront du retard.*

L'incertitude peut donc prendre deux formes : celle due à l'information incomplète (biais de communications, manque d'informations) et celle due à des aléas (environnement, capacités). Chacune de ces formes correspond à la levée d'une hypothèse présentée au chapitre précédent, respectivement à celle de connaissance a piori et à celle de déterminisme. La littérature a partiellement traité de la modélisation de ces formes d'incertitude dans la formation de coalitions, et nous allons voir comment dans cette section. En revanche, ces modèles sont souvent contextuels, et les deux formes d'incertitude ne sont que très rarement traitées ensemble. Les modèles ainsi proposés intègrent généralement une notion de répétition, dont nous développeront l'intérêt que les agents y trouvent. Cependant, la répétition de ces modèles n'est pas uniforme non plus, celle-ci pouvant être sur le jeu, sur le protocole ou bien par transition d'états. Afin de proposer par la suite un mo-

dèle stochastique avec un mécanisme de répétition simple, nous ferons une analogie entre la formation de coalitions répétée stochastique et les bandits manchots, un problème de décision séquentielle, ce qui nous permettra par ailleurs d'utiliser les stratégies de décision de ces derniers.

2.1.1 Formation de coalitions dans l'incertitude

Les différentes sources possibles d'incertitude peuvent donc être modélisées dans les jeux de coalitions. Pour cela, plusieurs façons sont envisageables, que nous pouvons regrouper en deux grands types de modèles : les jeux à informations privées et les jeux à fonctions caractéristiques stochastiques. Les premiers vont permettre de modéliser des situations où les agents ont par exemple des capacités qui ne sont pas nécessairement connues des autres agents.

Exemple 15. *Dans notre exemple du port maritime, les opérateurs sont nombreux et ont des compétences diverses qui ne sont pas connues de leurs collègues, comme par exemple la capacité à utiliser certains outils.*

Les jeux à fonctions caractéristiques stochastiques quant à eux vont permettre la modélisation de l'incertitude provenant de l'environnement ou bien de l'efficacité ponctuelle des agents.

Exemple 16. *Si nous reprenons les opérateurs, leur forme physique et mentale peut être variable, et malgré leurs capacités, ils peuvent parfois ne pas être en mesure de les utiliser avec autant de maîtrise que d'ordinaire.*

Ces deux types de modèles amènent de premiers éléments de réponse à la façon de lever les hypothèses de connaissance *a priori* et de déterminisme des utilités. En effet, l'information incomplète qu'ont les agents sur les autres dans les jeux à informations privées peut être vue comme une absence de connaissance *a priori*, tandis que pour les jeux à fonctions caractéristiques stochastiques, l'hypothèse levée est plus claire : si les utilités décrites par les fonctions caractéristiques sont stochastiques, alors l'hypothèse de déterminisme de ces utilités s'en retrouve levée. Intéressons-nous donc désormais de plus près à ces deux catégories de jeux de coalitions, mais également au mécanisme de répétition présent dans les modèles, qui permet aux agents de réduire leur incertitude, quelque soit la forme.

2.1.2 Jeux à informations privées

Une première source d'incertitude peut venir du fait que les agents disposent d'informations privées. Dans la littérature, ces informations privées décrivent les connaissances des agents sur leurs capacités ou celles des autres. Les modèles qui intègrent cette forme d'incertitude sont donc des extensions des jeux à capacités (voir 1.3.2.2), qu'ils soient quantitatifs [Blankenburg *et al.*, 2005, Shehory et Kraus, 1996, Shehory et Kraus, 1998, Shehory et Kraus, 1995] ou qualitatifs [Kraus *et al.*, 2003].

Dans un cadre quantitatif, les *jeux à informations privées* ajoutent aux jeux à capacités une *probabilité de succès* associée à chaque agent. Ceci représente une fiabilité dans la réalisation de tâches (et donc une capacité quantitative) [Blankenburg *et al.*, 2005, Shehory et Kraus, 1996, Shehory et Kraus, 1998, Shehory et Kraus, 1995]. Cette probabilité de succès est une information privée des agents : ils ne connaissent pas les probabilités de succès des autres agents. En ce qui les concerne, Shehory et Kraus [Shehory et Kraus, 1996, Shehory et Kraus, 1998, Shehory et Kraus, 1995] forment des coalitions pour des problèmes d'affectation de tâches. Ici, les agents possèdent des capacités quantitatives $S_i = \langle s_1^i, \ldots, s_r^i \rangle$, et doivent former des coalitions pour réaliser des tâches $T = \{t_1, \ldots, t_m\}$. Chaque tâche t_l requiert un ensemble de capacités $S_l = \langle s_1^l, \ldots, s_r^l \rangle$ pour être réalisées, où la somme des capacités des agents doit être supérieure aux capacités requises. Les agents ne connaissent donc que leur propre utilité et communiquent avec les autres agents pour calculer une utilité espérée pour chaque coalition. Le processus est le suivant : les agents créent une liste restreinte de coalitions de taille maximale fixée k qu'ils souhaitent former. Ils contactent les agents impliqués afin d'estimer leurs valeurs. Chaque agent a_i calcule localement une valeur w_S^i appelée *poids*, fondée sur l'utilité espérée, le coût de formation et la taille de chaque coalition C. Ensuite, la coalition avec le poids le plus faible est formée. Les solutions sont considérées comme stables tant que les coalitions formées peuvent réaliser leurs tâches, même si ce ne sont pas des solutions optimales.

Concernant Blankenburg *et al.* [Blankenburg *et al.*, 2005], ils proposent un modèle de jeu de coalitions sous la forme d'un protocole afin de déterminer quels agents sont dignes de confiance – c'est-à-dire fiables dans l'exécution de leur travail – pour former une coalition, également dans un problème d'allocation de tâches requérant des capacités que les agents peuvent avoir. Le protocole se déroule en quatre étapes : (1) communication des croyances sur les agents et des utilités observées des coalitions ; (2) formation des coalitions selon une variante du noyau ; (3) paiement des agents avant la réalisation de la tâche pour les inciter à adopter une bonne conduite ; (4) exécution de tâches et évaluation.

Plus formellement, pour l'agent a_i, l'étape (1) consiste à demander aux autres agents leur perception d'une probabilité de succès $\eta_i^j \in [0,1]$ concernant l'agent a_j, et à calculer ensuite $p_i^j = g(\{\eta_1^j, \ldots, \eta_i^j, \ldots, \eta_N^j\})$, c'est-à-dire la confiance que a_i a dans a_j avec une fonction d'agrégation g (par exemple une simple somme pondérée). Le concept de solutions ici consiste en l'adaptation du concept du noyau, afin de prendre en compte cette confiance des agents envers les autres.

Dans un cadre qualitatif, l'information privée ne s'applique plus aux capacités. En effet, dans un tel cadre, un agent sait exactement quelles compétences il possède et ces dernières ne sont pas quantifiées. C'est pourquoi dans le modèle de Kraus *et al.* [Kraus *et al.*, 2003], il est ajouté un coût b_i^t à chaque tâche t, coût qui est privé et spécifique à chaque agent, ici a_i. Kraus *et al.* réduisent alors le problème de formation de coalitions à un problème d'enchères descendantes, où le coût $B_C^t = \sum_{j \in C} b_j^t$ est celui de la coalition C pour la tâche t. Un commissaire-priseur propose des tâches à un prix p^t et les agents peuvent tous proposer des coalitions. Les membres de ces coalitions peuvent ou non accepter selon ce qu'ils estiment des coûts privés des autres agents. Si la coalition est formée, alors l'utilité de la coalition C pour la tâche t est $v(C,t) = p^t - B_C^t$. Il n'y a donc pas de concept de stabilité ici.

Notons que les modèles présentés ici présentent tous une perte d'optimalité, les agents prenant des décisions avec des éléments inconnus, et formant alors des coalitions dans lesquelles leurs gains sont moindres que ce qu'ils avaient escomptés. Notons que ces approches ont une autre conséquence : le processus de formation de coalitions doit nécessairement s'appuyer sur des éléments qui complexifient le modèle, tels que des contraintes de communications ou l'utilisation de mécanismes d'enchères.

2.1.3 Fonctions caractéristiques stochastiques

Dans un jeu classique, la fonction caractéristique qui décrit la valeur produite par une coalition est déterministe. Or, une seconde source d'incertitude peut être non plus les capacités des agents en elles-mêmes mais le résultat de l'exécution des tâches qu'entreprennent les coalitions. En effet, dans un contexte applicatif, il n'est pas garanti que les agents puissent connaître avec certitude le résultat de leurs actions, indépendamment de leurs compétences ou de leurs ressources. Ceci a conduit à définir des *jeux à fonctions caractéristiques stochastiques* [Blankenburg *et al.*, 2003, Chalkiadakis et Boutilier, 2004, Chalkiadakis et Boutilier, 2008, Chalkiadakis *et al.*, 2007, Charnes et Granot, 1976, Charnes et Granot, 1973, Gallardo et Jiménez-Losada, 2020, Ieong et Shoham,

2008].

Une première manière de modéliser la stochasticité consiste à associer à chaque coalition non pas une valeur mais une variable aléatoire bayésienne [Charnes et Granot, 1976, Charnes et Granot, 1973] ou floue [Blankenburg *et al.*, 2003, Gallardo et Jiménez-Losada, 2020, Mareš, 2001]. Par exemple dans le cadre bayésien, les travaux de Charnes et Granot définissent la fonction caractéristique comme $v : 2^N \rightarrow \mathcal{X}_{2^N}$ où les utilités des coalitions sont donc définies par des variables aléatoires suivant des lois normales [Charnes et Granot, 1973, Charnes et Granot, 1976]. Les imputations sont calculées à partir du gain espéré μ_C de chaque coalition C et les concepts de solutions sont redéfinis sur l'espérance de la fonction caractéristique, comme le *nucléole a priori* [Charnes et Granot, 1973].

Si cette approche est très générale, elle ne permet pas de modéliser les raisons sous-jacentes à l'incertitude. C'est pourquoi d'autres approches proposent d'augmenter les jeux de coalitions avec un modèle d'environnement stochastique, modélisant le fait que les agents ont une incertitude sur les effets de leurs actions. Par exemple, Chalkiadakis et Boutilier associent des jeux de coalitions à des processus de décision markoviens partiellement observables (POMDP) [Chalkiadakis et Boutilier, 2008]. Ici, les agents possèdent chacun un type spécifique $t_i \in T_i$ où T_i est l'ensemble des types spécifiques possibles pour l'agent a_i, un type décrivant un ensemble de capacités lui permettant de réaliser des *actions coalitionnelles* (c'est-à-dire effectuer une action au sein d'une coalition). Chaque agent a_i connait son type t_i mais ne connaît pas le type t_j de chaque autre agent a_j. Il a en revanche une croyance $B_i(\vec{t}_{N\setminus a_i})$ sur l'ensemble joint des types possibles de tous les autres agents $\vec{t}_{N\setminus a_i} = \{t \in T_j, \forall a_j \in N \setminus a_i\}$. Bien que nous sommes en présence d'informations privées, une autre source d'incertitude est présente : lorsqu'une coalition C effectue une action α, il en résulte un état $s \in S$ (S étant l'ensemble des états possibles dans le jeu) avec une certaine probabilité $Pr(s|\alpha, \vec{t}_C)$, dépendant de l'action α et des types réels des agents de la coalition $\vec{t}_C$. Cependant, en raison de la fonction de transition non-déterministe du POMDP, une même action coalitionnelle par une coalition C (et donc les mêmes types réels) peut amener à plusieurs états s différents. Enfin, atteindre un état s produit un gain r_s qui peut être distribué entre les agents de la coalition ayant réalisé l'action amenant à cet état. Le gain obtenu suite à une action est donc stochastique, c'est pourquoi les agents formulent un vecteur de demande $\vec{d} = \langle d_1, \ldots, d_n \rangle$ décrivant le gain qu'ils estiment suite à l'action coalitionnelle qu'ils envisagent. Celui-ci a pour objectif de remplacer le vecteur de gain $\vec{x}$ dans la solution. Le vecteur peut être réduit à une coalition tel que $\vec{d}_C$ est le vecteur de demande pour les agents de C. Cela permet de faire une proportionali-

sation sur le gain réel suite à l'action coalitionnelle effectuée. Afin de formaliser la notion de stabilité, Chalkiadakis et Boutilier définissent le *cœur bayésien* [Chalkiadakis et Boutilier, 2004, Chalkiadakis *et al.*, 2007], qui est semblable à l'approche de Charnes et Granot, c'est-à-dire en s'appuyant sur l'espérance des valeurs des coalitions en fonction des actions jointes de leurs membres. Pour cela, l'agent a_i définit une estimation du gain $\vec{p}^{\,i}_{j,C}(\alpha, \vec{d}_C)$ de l'agent a_j si celui-ci était membre de la coalition C avec le vecteur de demande $\vec{d}_C$. Le cœur bayésien est donc l'ensemble des solutions $S = \langle \mathcal{CS}, \vec{d} \rangle$ pour lesquelles, pour toute coalition $C \in \mathcal{CS}$, il existe une action α telle que il n'existe aucune coalition C', aucune action β et aucun vecteur de demande $\vec{d}_{C'}$ pour lesquels $\vec{p}^{\,i}_{i,C}(\beta, \vec{d}_{C'}) > \vec{p}^{\,i}_{i,C}(\alpha, \vec{d}_C)$.

Dans le même ordre d'idée, Ieong et Shoham ont proposé des jeux de coalitions associés à des modèles de mondes possibles [Ieong et Shoham, 2008]. Ici, les agents représentent leur incertitude comme une distribution de probabilités sur un ensemble de mondes $\Omega = \{\omega_1, \ldots, \omega_m\}$, chacun représentant un jeu de coalitions avec une fonction caractéristique déterministe unique $v^{\omega_k} : 2^N \to \mathbb{R}$. Une distribution de probabilité $\mathbb{P}$ sur l'existence de ces mondes est connue et commune à tous les agents, cependant chaque agent a_j possède un sous-ensemble $\mathcal{I}_j \subseteq \Omega$ qui réduit sa croyance d'existence des mondes à un sous-ensemble de Ω. À chaque monde ω_k est associé un ensemble fini d'imputations sur lesquelles chaque agent a_j exprime des préférences $\succcurlyeq_i (\omega_k)$. Ieong et Shoham proposent alors de nouveaux concepts de solutions fondés, certes sur les préférences établies par les agents, mais surtout sur une notion de connaissance spécifique. Le concept *ex-ante* caractérise une solution avant l'observation du monde véritable, le concept *ex-interim* après réduction des mondes possibles lors que les agents reçoivent des informations mais avant l'observation du monde véritable, et le concept *ex-post* après l'observation du monde véritable. Sans incertitude, ces trois concepts équivalent au cœur.

Nous avons donc deux grandes catégories d'incertitude au sein de la formation de coalitions : celle où l'information est incomplète et celle où les utilités sont stochastiques. Dans certains des travaux présentés, le modèle d'incertitude est connu des agents, et ceux-ci se fondent dessus afin de prendre leur décision, mais lorsque ce n'est pas le cas, les agents doivent estimer ce modèle. De manière intéressante, lorsqu'un jeu à information incomplète ou un jeu stochastique est répété, les agents peuvent tirer parti de cette répétition pour raffiner leurs estimations.

2.1.4 Incertitude et répétition

Lorsque nous souhaitons modéliser des cadres applicatifs réels contenant de l'incertitude, ceux-ci sont souvent dynamiques. En effet, si nous prenons en exemple une incertitude environnementale comme les aléas météorologiques, il ne semble pas pertinent d'en tenir compte dans un système statique. Néanmoins, le fait de se placer dans un cadre dynamique pour intégrer de manière plus naturelle l'incertitude permet aux agents d'en profiter pour tenter de réduire cette incertitude.

En section 1.2, nous avons présenté divers types de jeu proposés par la théorie des jeux. Certains d'entre eux, comme les jeux simultanés, peuvent être répétés, à l'instar du dilemme du prisonnier itéré dans les travaux d'Axelrod [Axelrod et Hamilton, 1981]. Si son centre d'intérêt était l'évolution des stratégies dans le temps, le nôtre est plus subtil. La répétition d'un jeu permet entre autres aux agents d'observer l'état du jeu à chaque itération. C'est donc particulièrement cette capacité d'observation qui nous intéresse : afin de permettre une meilleure décision dans un cadre stochastique, l'observation est essentielle pour réduire l'incertitude quelque soit sa forme. En estimant avec davantage de précision les différentes utilités des coalitions, les agents peuvent déterminer quelles coalitions former en priorité. La formation de coalitions étant un problème se modélisant par des jeux, nous pouvons donc appliquer le même procédé, c'est-à-dire que le problème est répété dans le temps. Nous retrouvons alors des modèles utilisant ce mécanisme de répétition dans la littérature.

Concernant les jeux à informations privées, il existe deux approches que nous avons déjà cité qui sont fondées sur la répétition des jeux, pour tenir compte de l'incertitude qui en résulte. D'un côté, dans les travaux de Blankenburg *et al.* [Blankenburg *et al.*, 2005], la répétition du jeu consiste à exécuter le protocole plusieurs fois, formant diverses coalitions et en évaluant le résultat *a posteriori*. Ainsi, l'agent a_i apprend la fiabilité des autres agents a_j, c'est-à-dire leur probabilité de succès η_i^j estimée, en observant les utilités des coalitions auxquelles ils participent et l'exécution des tâches. Grâce à cette estimation de la fiabilité, les agents peuvent donc à nouveau agréger les croyances de tous les agents pour mettre à jour $p_i^j = g(\{\eta_1^j, \ldots, \eta_i^j, \ldots, \eta_N^j\})$, c'est-à-dire la confiance que l'agent a_i a dans a_j, sur le même principe que les systèmes de réputation [Sabater et Sierra, 2005]. Cela permet donc aux agents de potentiellement effectuer un choix différent de coalition à la répétition suivante. D'un autre côté, dans les modèles de Shehory et Kraus [Shehory et Kraus, 1995, Shehory et Kraus, 1996, Shehory et Kraus, 1998], c'est le jeu lui-même qui est répété. Pour rappel, les agents ne connaissent que leur propre utilité, en raison de

leurs capacités $S_i = \langle s_1^i, \ldots, s_r^i \rangle$, et demandent explicitement à chaque autre agent a_j leurs capacités S_j, si ceux-ci font partie d'une liste restreinte de coalitions de taille maximale k. Cependant, ces capacités S_j peuvent changer au cours du temps, les agents doivent donc régulièrement répéter l'étape de communication afin de mettre à jour les *poids* w_C^i estimés pour chaque coalition C. Il y a donc un fort coût de communication bien que cela réduise l'incertitude.

Pour les jeux à fonction caractéristique stochastique, nous pouvons citer à nouveau Chalkiadakis et Boutilier [Chalkiadakis et Boutilier, 2008] dont le modèle repose sur un processus d'apprentissage par renforcement bayésien sur un POMDP, qui peut être vu comme un jeu répété. Concernant l'apprentissage par renforcement, ce dernier utilise les observations pour mettre à jour le modèle estimé et l'utiliser tout de suite après. L'algorithme décide alors d'une action qui va influer sur l'environnement, ce qui va alors donner des nouvelles observations à l'algorithme, et ainsi de suite.

Exemple 17. *Un exemple sur le port maritime est quand les douaniers examinent la marchandise. Au départ, ils peuvent choisir aléatoirement quels conteneurs ils vont examiner, mais s'ils se rendent compte de problèmes concernant une compagnie en particulier (l'observation), ils vont alors se concentrer sur leurs conteneurs (l'action). S'il se trouve que c'était un seul conteneur qui posait problème (une nouvelle observation), alors ils vont reprendre leur analyse aléatoire (une nouvelle action).*

Dans le modèle de Chalkiadakis et Boutilier, les agents doivent donc choisir à chaque itération du POMDP de former une coalition C. Une fois les coalitions formées, ces dernières réalisent une action α qui provoque une transition stochastique vers un état du monde s avec la probabilité $Pr(s|\alpha, \vec{t}_C)$. Chaque agent observe le gain r_s produit dans ce nouvel état et met à jour son état de croyance $B_i(\vec{t}_{N \setminus a_i})$. Cela leur permet d'apprendre d'une part les capacités des autres agents mais également le modèle de transition stochastique entre les états. Ils peuvent donc ensuite produire une meilleure estimation $\vec{p}_{j,C}^{\,i}(\alpha, \vec{d}_C)$ concernant les gains de chaque agent a_j pour la formation de la coalition C, et de produire un meilleur vecteur de demande $\vec{d}$ (qui représente l'espérance de gain de tous les agents).

Nous avons donc vu différentes manières de modéliser l'incertitude, qu'elle soit due à un manque d'informations ou à une stochasticité des utilités elles-mêmes, mais également comment profiter de la répétition d'un jeu pour gérer cette incertitude. Regroupons l'ensemble de ces éléments dans une catégorie précise : la formation de coalitions stochastique répétée. Cependant, ces gains incertains et la répétition du jeu nous renvoient à un autre problème de décision : celui des bandits manchots.

2.2 Bandits manchots et formation de coalitions stochastique répétée

Le problème des bandits manchots (ou MAB) est un problème de décision séquentielle avec apprentissage qui présente des similitudes avec la formation de coalitions stochastique répétée. Nous nous permettons alors de faire un analogie entre les deux problèmes, bien que des limites à celle-ci existent. De plus, nous présentons dans cette section des stratégies des bandits manchots, qui utilisent un mécanisme spécifique permettant de lier la maximisation du gain et la gestion de l'incertitude.

2.2.1 Analogie avec les bandits manchots

Une analogie de la formation de coalitions stochastique répétée peut être faite avec le problème des bandits manchots. En effet dans ce problème, un agent doit choisir de tirer sur un bras parmi d'autres, ce qui va générer un gain stochastique, sans que l'agent connaisse le modèle stochastique derrière le bras. Nous allons donc présenter ce problème puis développer l'analogie.

2.2.1.1 Un problème de décision séquentielle

Le problème des bandits manchots est donc un problème de décision séquentielle [Gittins, 1979]. Celui-ci se compose d'un ensemble de bras qui, lorsqu'ils sont tirés (ou actionnés) par un agent, rapportent un gain. Ce gain est aléatoire selon une distribution de probabilité cachée de l'agent. Au début de la séquence, l'agent a donc une ignorance totale des distributions de probabilité des bras, et profite des observations du gain donné par les différents bras faites à chaque pas de temps pour tenter d'estimer ces distributions.

Définition 2.1 (Bandits manchots)**.** *Soient :*

- *a un agent,*
- π_a *la* politique de décision *de l'agent a,*
- $\mathbb{T} = \{t_1, \ldots, t_n\}$ *un ensemble de pas de temps de taille n,*
- $B = \{b_1, \ldots, b_k\}$ *un ensemble de bras de taille k,*
- $\Phi = \{\phi_1, \ldots, \phi_k\}$ *un ensemble de distributions de probabilité,*
- $\Psi = \{\psi_1, \ldots, \psi_k\}$ *un ensemble de distributions de probabilité estimées.*

À chaque pas de temps $t_m \in \mathbb{T}$, l'agent a utilise π_a pour déterminer le bras $b_j \in B$ à tirer. Un gain aléatoire suivant la distribution ϕ_j est alors obtenu, et l'agent met à jour

sa croyance ψ_j concernant la distribution de probabilité du bras.

Afin de construire l'ensemble de distributions de probabilité estimées, l'agent s'appuie sur les observations qu'il peut faire à chaque pas de temps.

Définition 2.2 (Observations)**.** *Soit $\mathcal{O}_t = \{(b_j, t', r_j^{t'})) : t' \in \mathbb{T}, t' < t\}$ un ensemble d'observations au pas de temps t correspondant à l'ensemble des bras tirés à chaque pas de temps avant t et leurs gains réels produits.*

Selon le type de lois de probabilité utilisé, le calcul de l'estimation peut être différent, mais s'il s'agit de lois normales, l'estimation peut être réalisée en calculant la moyenne et la variance des gains observés pour chaque bras. Concernant la politique de décision, qui détermine quel bras l'agent choisira de tirer, utilise un mécanisme spécifique, à savoir l'*équilibre exploration-exploitation*, qui sera présenté dans la suite, après que l'analogie avec la formation de coalitions stochastique répétée ait été développée.

2.2.1.2 Liens avec la formation de coalitions stochastique répétée

Des liens peuvent être établis entre les bandits manchots et un problème de formation de coalitions stochastique répétée. Les éléments de comparaison qui nous seront utiles sont présentés par le tableau 2.1.

	Bandits Manchots	**Formation de coalitions**
Points communs	Bras	Coalitions
	Politique de décision	Concept de solutions
	Gains stochastiques	Utilités des coalitions
Différences	1 bras tiré	1 à N coalitions formées
	Bras indépendants	Fonction caractéristique structurée

TABLE 2.1 – Comparaison entre les bandits manchots et la formation de coalitions

La formation de coalitions étant traditionnellement un problème centralisé, nous pouvons faire le rapprochement entre l'agent du bandit manchot et le décisionnaire central de la formation de coalitions. Former une coalition revient donc à tirer un bras : cela produit un gain (ou une utilité) qui peut être observé. Dans le cadre de la formation de coalitions stochastique telle que l'ont décrit Charnes et Granot [Charnes et Granot, 1973], les utilités sont décrites par des variables aléatoires suivant des lois de probabilité, tout comme dans les bandits manchots. Pour finir, si ce qui définit le choix du bras à tirer dans les

bandits manchots est la politique de décision, et que les bras peuvent être vus comme des coalitions, alors les concepts de solutions de la formation de coalitions décrivant quelles coalitions sont stables sont équivalent à la politique de décision.

Cependant, nous n'avons pas encore présenté les politiques de décision, ce que nous allons faire dans la section suivante.

2.2.2 Équilibre exploration-exploitation

Le problème des bandits manchots a été décrit comme faisant partie de la famille des problèmes d'apprentissage par renforcement. En effet, les politiques de décision utilisées dans ce problème sont des algorithmes de ce type d'apprentissage, et se font en ligne. Cependant, ces algorithmes font appel à un mécanisme particulier de l'apprentissage par renforcement : l'*équilibre exploration-exploitation*.

L'*équilibre exploration-exploitation* est un mécanisme utilisé dans des cadres où l'incertitude est présente, qui permet d'un côté la maximisation du gain dans le temps (exploitation) et, de l'autre, de réduire l'incertitude par l'observation (exploration). Il existe plusieurs stratégies s'appuyant sur cet équilibre. Nous en présentons ici une revue non-exhaustive, à savoir les stratégies *ϵ-gloutonne* [Tran-Thanh *et al.*, 2010], *UCB* [Agrawal, 1995, Auer *et al.*, 2002], et *EXP3* [Auer *et al.*, 2002]. La première est dite *non contextuelle* car elle choisit un bras uniquement en fonction de son espérance, et accepte parfois de ne pas choisir le meilleur afin de gagner en information, tandis que les stratégies *contextuelles* (UCB et EXP3 donc) se fondent sur un compromis entre l'exploration et l'exploitation [Vallée, 2015].

2.2.2.1 Stratégie ϵ-gloutonne

Une première stratégie s'appuyant sur un mécanisme d'exploration-exploitation est la stratégie ϵ-gloutonne [Tran-Thanh *et al.*, 2010]. Le principe de celle-ci est simple : avec une certaine probabilité ϵ donnée, l'algorithme va explorer, c'est-à-dire effectuer une action aléatoire parmi toutes celles possibles. Avec la probabilité opposée, soit $1 - \epsilon$, l'algorithme va effectuer l'action avec laquelle le gain sera maximisé selon l'état actuel des connaissances de l'agent. Nous sommes donc en présence d'une approche gloutonne qui cherche à maximiser le gain espéré mais qui accepte de temps en temps de gagner moins, au profit d'un gain aléatoire de connaissances. Formellement, nous pouvons écrire cette stratégie comme suit.

Définition 2.3 (ϵ-glouton)**.** *Le bras b_i^{t+1} qui sera tiré au pas de temps $t+1$ est donné par l'algorithme suivant :*

1. *Tirer un nombre aléatoire γ entre 0 et 1,*
2. *Si $\gamma \leq \epsilon$, alors tirer un bras $b_i \in B$ aléatoirement,*
3. *Sinon, $b_i^{t+1} = \text{argmax}_{b_j \in B}\, r_j$ où r_j est l'utilité espérée du bras b_j.*

2.2.2.2 Stratégie UCB

La stratégie UCB (pour *Upper Confidence Bound*) est également une stratégie fondée sur un mécanisme d'exploration-exploitation [Agrawal, 1995, Auer *et al.*, 2002]. Dans celle-ci, il n'est pas question de probabilité mais uniquement d'une agrégation de valeurs entre le gain espéré et un biais d'exploration caractérisant l'absence d'information quant à un élément. En effet, en prenant en compte le nombre total d'interactions (c'est-à-dire combien de fois un bras quelconque a été tiré) et le nombre d'interactions avec un bras précis, cette formule produit un biais d'exploration. Le poids de ce biais dans la stratégie UCB croîtra en même temps que le nombre total d'interactions si ce bras précis n'est pas celui tiré, et décroîtra si c'est celui tiré.

Définition 2.4 (UCB)**.** *Le bras b_i^{t+1} qui sera tiré au pas de temps $t+1$ est donné par :*

$$b_i^{t+1} = \underset{b_j \in B}{\text{argmax}}\, r_j + \sqrt{\frac{2ln(t)}{n_j}},$$

où r_j est l'utilité espérée du bras b_j, et n_j le nombre d'observations de ce même bras.

2.2.2.3 Stratégie EXP3

La stratégie EXP3 (pour *Exponential-weight algorithm for Exploration and Exploitation*) est un autre stratégie bien connue dans les bandits manchots [Auer *et al.*, 2002]. Cette stratégie repose essentiellement sur l'apprentissage de poids afin de définir quels sont les bras les plus prometteurs au fur et à mesure que le problème se déroule.

Définition 2.5 (EXP3)**.** *Soient γ un paramètre d'exploration et $B = \{b_1, \ldots, b_k\}$ un ensemble de bras donnés. Chaque bras $b_i \in B$ possède un poids w_{b_i} initialisé à 1. Le bras b_i^t qui est tiré au pas de temps t est donné par l'algorithme suivant :*

1. *Initialiser $p_i(t) = (1-\gamma)\frac{w_{b_i}(t)}{\sum_{j=1}^{k} w_{b_j}(t)} + \frac{\gamma}{k}$ pour chaque bras $b_i \in B$,*

2. *Tirer aléatoirement le prochain bras* b_i^t *selon les probabilités* $p_i(t), \ldots, p_k(t)$,
3. *Observer la récompense* $x_{b_i^t}(t) \in [0, 1]$,
4. *Définir la récompense estimée* $\hat{x}_{b_j^t}$ *tel que :* $\frac{x_{b_j^t}(t)}{p_{b_j^t}(t)}$ *pour* $b_j^t = b_i^t$, 0 *pour tous les autres* $b_j^t \in B \setminus \{b_i^t\}$,
5. *Mettre à jour le poids tel quel :* $w_{b_i^t}(t+1) = w_{b_i^t}(t) e^{\gamma \hat{x}_{b_i^t}(t)/K}$.

Cependant, cette stratégie a été proposée dans le cadre des bandits manchots *truqués*, où aucune hypothèse stochastique n'est faite sur la génération des gains des bras, du fait qu'un *adversaire* décide à chaque pas de temps du gain des bras [Auer *et al.*, 1995].

Ces stratégies permettent donc aux agents de minimiser leur regret en choisissant parfois d'explorer, mais sans perdre de vue le gain. Cependant, ces stratégies font l'hypothèse que les bras sont dépendants, ce qui peut amener certaines limites dans l'analogie avec la formation de coalitions.

2.2.3 Dépendance et apprentissage

Les stratégies présentées ci-dessus ont donc toutes un fonctionnement similaire, avec une exploration aléatoire et une exploitation gloutonne, ainsi qu'un apprentissage en ligne qui ne considère aucune dépendance. Considérer les bras du problème du bandit manchot comme étant indépendants les uns des autres est une hypothèse courante et qui se tient, cependant, cela n'est pas nécessairement souhaitable lorsque l'on parle de coalitions.

2.2.3.1 Limites de cet équilibre avec l'analogie

Contrairement aux problèmes classiques de bandits manchots où les bras sont indépendants, les fonctions caractéristiques de la formation de coalitions tendent à posséder une structure (voir 1.3.2). Ces structures caractérisent une synergie sous-jacente entre les agents et des corrélations sur les tailles des coalitions. Une question qui se pose donc est celle de l'inférence : pouvons-nous déduire, à partir d'une observation concernant une coalition, des informations sur d'autres coalitions de même taille, ou possédant en partie les mêmes agents ? Ce mécanisme logique permet, en apprentissage, de déduire des informations sur certains éléments à partir des données ou observations d'éléments proches. L'inférence permet notamment d'accélérer l'apprentissage de modèles structurés ou contenant des synergies.

Exemple 18. *Un contremaître sur un port maritime pourra déduire, suite à l'observation d'un dockeur très efficace avec deux autres dockeurs précis individuellement, que rassembler les trois dockeurs pourrait être encore davantage efficace.*

Les méthodes d'apprentissage par renforcement précédemment présentées ne tiennent pas compte de possibles dépendances, il peut donc être intéressant de s'intéresser à d'autres formes d'apprentissage qui intégreraient cette notion d'inférence. Une autre limite, bien que moins contraignante, est celle du nombre de bras tirés simultanément. En effet, dans le problème des bandits manchots, l'agent tire un bras à chaque fois, tandis que dans un cadre de formation de coalitions, si un bras représente une coalition, le nombre de bras à tirer simultanément est compris entre un (la grande coalition) et le nombre d'agents n (l'ensemble des coalitions singleton), parmi 2^n bras. Outre le fait que le nombre de bras à tirer est variable selon la cardinalité de la structure de coalitions (en 1 et n), il faut pouvoir respecter la contrainte des coalitions disjointes, c'est-à-dire que certains bras ne doivent pas être tirés ensemble.

2.2.3.2 Les super-bras dans les bandits manchots

Pour commencer, Anantharam *et al.* [Anantharam *et al.*, 1987] ont proposé une variante de bandits manchots où l'agent devait tirer m bras à chaque pas de temps, parmi les k bras du problème. Cependant, plusieurs contraintes pour la modélisation de la formation de coalitions via des bandits manchots ne sont pas respectées avec ce modèle. En effet, contrairement aux bandits manchots classique, les gains dans ce modèle sont non-markoviens, c'est-à-dire que le gain d'un bras à un pas de temps donné dépend de l'historique de ce même bras, c'est-à-dire s'il a été joué ou non. De plus, si les coalitions sont représentées par les bras, il est nécessaire de tirer des bras représentant des coalitions disjointes, ce qui n'est pas garanti dans ce modèle, l'agent pouvant tirer les m bras qu'il souhaite sans contrainte particulière hormis le nombre. Pour finir, les gains des bras sont totalement indépendants des autres et observés individuellement, ce qui n'implique aucune dépendance.

La littérature étudie néanmoins certains cadres de bandits manchots où de la dépendance est créée entre les bras du bandit : les *bandits combinatoires*. Dans ce cadre, les modèles intègrent des *super-bras* [Chen *et al.*, 2013]. Ces derniers permettent notamment de tirer simultanément un ensemble *fixe* de plusieurs bras, appelés *bras basiques*, ce qui induit une dépendance entre ces derniers. En effet, si le gain de chaque bras relié au même super-bras sera bien indépendant du gain des autres, le gain produit par

le super-bras dépend de tous ces bras. Selon les modèles de bandits combinatoires, seul le gain du super-bras peut-être observé (type *bandit*, et auquel cas la fonction de composition des gains est inconnue) ou les gains de tous les bras basiques est observé (type *semi-bandit*) [Audibert *et al.*, 2011].

Définition 2.6 (Bandits manchots combinatoires)**.** *Un bandit manchot combinatoire est un problème de bandit manchot (définition 2.1) auquel est ajouté un ensemble de super-bras $\mathcal{S} = \{S_1, \ldots, S_m\}$ où $S_n \subseteq B$. Tirer sur le super-bras S_n tire sur l'ensemble des bras $b_j \in S_n$.*

Des travaux ont aussi été menés sur un cadre spécifique des bandits combinatoires, qui ajoute une nouvelle couche d'incertitude. Lorsqu'un super-bras S est tiré, les bras basiques $B_S = \{b_1, b_2, b_3\}$ reliés à lui ne sont tirés qu'avec une certaine probabilité p_x indépendante entre ces bras, tel que $p_x \in \{p_1, p_2, p_3\}$ (où p_x est la probabilité pour le bras b_x). Si un bras basique b_x est finalement tiré via son super-bras S, alors il produit un gain stochastique selon la distribution de probabilité ϕ_x (définition 2.1) [Chen *et al.*, 2016]. Cet ajout d'incertitude rend néanmoins l'estimation des distributions de probabilité sous-jacentes des bras basiques plus difficile, d'autant plus que dans ce modèle, l'agent n'a pas de connaissances précises sur les ensembles de bras basiques reliés aux super-bras : il ne peut donc pas déterminer si un même bras basique fait partie de deux super-bras distincts.

Malgré la résolution du problème du tirage simultané d'un nombre variable de bras, modéliser le problème de formation de coalitions à l'aide de bandits combinatoire semble d'un intérêt limité. En effet, si pour un petit nombre d'agents, le nombre de super-bras serait inférieur au nombre de bras basiques, lorsque l'on dépasse $n = 5$, il y a davantage de structures de coalitions possibles que de coalitions possibles. De plus, si les bandits combinatoires intègrent une notion de dépendance, celle-ci ne découle que du bruit intrinsèque dû au gain unique au tirage d'un super-bras tandis qu'il est potentiellement créé par plusieurs bras basiques, ce qui ne résoud pas le problème d'indépendance structurelle des coalitions, et donc ne permet pas de déduire des synergies entre agents.

2.2.3.3 Inférence : réseaux de neurones

Un autre moyen d'explorer les dépendances possibles est l'inférence, c'est-à-dire la déduction de connaissances à propos d'un élément, par l'observation d'autres éléments. À nouveau, nous pouvons nous appuyer sur l'analogie entre la formation de coalitions

stochastique répétée et les bandits manchots afin d'explorer les possibilités de modélisation qui s'offrent à nous, et notamment concernant les méthodes d'apprentissage usant de l'inférence appliquées aux bandits manchots.

Parmi les modèles d'apprentissage utilisant l'inférence, les *réseaux de neurones* sont sans doute les plus connus à l'heure actuelle. Ces modèles du domaine de *l'apprentissage automatique* mettent en scène des *neurones artificiels* connectés au sein d'un réseau, qui s'inspirent schématiquement des neurones du cerveau humain. Bien qu'il existe plusieurs architectures, les plus connus (dits à *propagation directe*) sont organisés en couches successives de neurones : une couche d'entrée qui reçoit les données, un nombre variable de couches cachées contenant des neurones (là encore, leur nombre est variable) et une couche de sortie contenant un ou plusieurs neurones. Les réseaux de neurones servent à de nombreuses applications, comme par exemple la reconnaissance d'images, la classification de textes ou d'images ou encore la prédiction de données. À cette fin, les réseaux de neurones sont classiquement entraînés sur des jeux de données afin de déterminer le poids de chaque neurone et former une matrice de poids pour chaque couche cachée, en mesurant l'erreur entre l'objet test et le résultat que le réseau retourne, puis par une méthode de *descente de gradient*, corrige les poids des neurones pour minimiser l'erreur.

Des travaux de la littérature se sont intéressés à l'utilisation de réseaux de neurones pour les bandits manchots. Nous pouvons citer notamment Dawson *et al.* [Dawson *et al.*, 2009] ainsi que Collier et Llorens [Collier et Llorens, 2018]. Bien que le but des travaux de Dawson *et al.* est assez spécifique aux réseaux de neurones et leurs fonctions d'apprentissage, ils montrent que le problème des bandits manchots peut être modélisé à partir d'un réseau de neurones assez simple, nommé perceptron multicouches, c'est-à-dire un modèle à propagation directe. Ici, la couche d'entrée du réseau représente les bras du bandit, et l'activation d'un neurone d'entrée correspond au fait de tirer un bras. Le gain des bras est cependant restreint à une valeur binaire : soit il produit un gain de 0, soit de 1, avec respectivement une probabilité de p et $1 - p$. Les distributions de probabilité des bras de ce modèle suivent donc une *loi de Bernoulli*. Selon la valeur retournée par le réseau et une valeur aléatoire comprise entre 0 et 1 (tirée à chaque fois que le réseau doit donner une réponse), les poids des neurones du réseau sont mis à jour. Si la valeur aléatoire est plus faible que la valeur retournée par le réseau, alors le réseau applique une méthode de descente de gradient, sinon rien ne se passe, la mise à jour des poids n'est donc pas systématique. Enfin, la stratégie de bandits manchots utilisée est ϵ-gloutonne. Les auteurs concluent donc que les bandits manchots sont tout à fait modélisable par

des réseaux de neurones, et que contrairement aux travaux classiques sur les bandits manchots, l'estimation des gains des bras peut dépendre des autres bras, et ne sont pas totalement indépendants, en prenant en exemple qu'il peut exister des variables cachées possiblement communes à certains bras.

Concernant Collier et Llorens, ils utilisent les réseaux de neurones dans un cadre particulier des bandits manchots : les *bandits contextuels*. Dans ces derniers, un contexte est observé par l'agent avant de tirer un bras, et selon le contexte, les gains produits par les bras ne seront pas les mêmes [Auer *et al.*, 2002]. Tout comme Dawson *et al.*, Collier et Llorens modélisent le problème des bandits manchots grâce à un réseau de neurones, en incluant cette fois-ci le contexte. Le modèle de réseau de neurones est également semblable, avec un perceptron multicouches possédant deux couches cachées. Un autre point commun de ces travaux est la modélisation des distributions de probabilité régissant les gains des bras : ce sont des lois de Bernoulli (c'est-à-dire que le gain peut prendre les valeurs 0 ou 1 respectivement avec une probabilité p et $1 - p$). Les données d'entrée du réseau de neurones ne sont plus uniquement des actions (tirer un bras), mais un couple (x, α_i) où x est le contexte observé par l'agent, et α_i l'action de tirer le bras b_i. L'ensemble de ces couples sont donnés au réseau afin d'effectuer de l'inférence, et l'action α_j choisie par l'agent sera celle pour laquelle le couple (x, α_j) a le gain espéré maximal. Grâce aux poids du réseau, les auteurs modélisent une stratégie des bandits manchots fondée sur l'équilibre exploration-exploitation, à savoir l'*échantillonage de Thompson*, dont le principe repose sur la sélection de bras en fonction de leur gain espérée et d'un poids, mis à jour a posteriori. Ce poids est donc ici modélisé par le poids des neurones. Concernant la mise à jour de ces derniers, celle-ci se fait à des pas de temps précis définis par un variable *nextRetrain*, dont la valeur croît exponentiellement, pour laquelle est défini un coefficient K. Au début du processus, *nextRetrain* est équivalente au nombre de bras, puis à chaque mise à jour, la variable est multipliée par le coefficient K. Cela a pour conséquence d'entraîner souvent le réseau au début, quand les données sont peu nombreuses, et plus celles-ci se cumulent, moins le réseau est entraîné souvent. Les auteurs proposent également de remplacer l'échantillonage de Thompson par la stratégie ϵ-gloutonne et de comparer les deux versions à un bandit manchot contextuel modélisé classiquement. Leur conclusion est que leur modèle est empiriquement plus efficace qu'un bandit manchot contextuel modélisé classiquement, et que la version avec l'échantillonage de Thompson est plus performante. Cependant, cette stratégie étant décrite et étudiée pour les bandits contextuels, celle-ci est n'est pas d'un grand intérêt pour nous.

Il est donc possible de faire de l'inférence dans les bandits manchots, et cela sans utiliser des mécanismes particuliers comme les super-bras, en utilisant une méthode d'apprentissage telle que les réseaux de neurones, qui s'accordent bien avec les stratégies classiques et les fonctions de gains non-linéaires des bandits manchots.

2.3 Conclusion

Nous avons donc étudié ici les éléments sur lesquels nous pourrons nous appuyer afin de répondre à certaines de nos problématiques, à savoir la levée des hypothèses de déterminisme et de connaissance *a priori* des utilités des coalitions. En effet, certains modèles de la littérature ont déjà proposé des modèles où l'utilité des coalitions est stochastique via des fonctions caractéristiques stochastiques, ou des modèles où l'information des agents à propos des autres est incomplète avec les jeux de coalitions à informations privées. Cependant, ces modèles sont en général très contextuels, c'est pourquoi afin de répondre à nos problématiques, nous souhaitons proposer un nouveau modèle de formation de coalitions stochastique répétée, et utiliser des stratégies des bandits manchots fondées sur l'équilibre exploration-exploitation. Ce nouveau modèle devant être le moins contextuel possible, celui-ci sera fondé sur celui de Charnes et Granot [Charnes et Granot, 1973, Charnes et Granot, 1976], dans lequel les utilités sont représentées par des variables aléatoires. En effet, les informations privées de type capacités ne sont pas notre centre d'intérêt principal, de plus, si nous voulons les intégrer, il est tout à fait possible d'intégrer l'incertitude due à ces capacités au sein des distributions de probabilité régissant les variables aléatoires du modèle de Charnes et Granot. Cependant, la nature privée des informations est intéressante pour lever l'hypothèse de connaissance *a priori*. Les modèles ont montré qu'il est possible de raisonner sur des éléments inconnus *a priori*, et nous allons nous inspirer de cela pour fusionner les deux types d'incertitude en un seul modèle : les utilités seront stochastiques, mais également inconnues des agents.

Il a aussi été montré que dans des problèmes de formation de coalitions contenant de l'incertitude, les modèles intègrent souvent une notion temporelle ou de répétition, ce qui permet aux agents d'observer divers éléments comme les utilités des coalitions ou la fiabilité des autres agents, afin de réduire l'incertitude et de prendre la meilleure décision. Ce mécanisme de répétition et d'observation nous a par ailleurs permis de faire une analogie avec le problème de décision séquentielle des bandits manchots, analogie qui nous sera utile par la suite pour proposer ce nouveau modèle de formation de coalitions

stochastique répétée indépendant de tout contexte. C'est notamment par le prisme de cette analogie que nous allons aborder le problème de la formation de coalitions stochastique répétée, en particulier avec l'utilisation de stratégies classiques des bandits manchots. Un autre élément d'intérêt est la possibilité de modélisation du problème des bandits manchots par des réseaux de neurones, et donc par extension, la formation de coalitions également. Cela permettra d'effectuer de l'inférence sur les utilités des coalitions, en raison de la structuration possible des fonctions caractéristiques.

Chapitre 3

Formation de coalitions stochastique répétée

Sommaire

Comme montré dans le chapitre 2, la littérature a étudié des questions connexes aux nôtres, notamment sur l'incertitude dans les jeux de coalitions. Cependant, les modèles proposés sont souvent contraints, tandis que nous souhaitons lever les hypothèses de déterminisme des utilités et de connaissance *a priori* de ces dernières dans le cadre le plus général possible. C'est pourquoi nous présentons dans ce chapitre un nouveau modèle de jeux de coalitions dans un cadre stochastique et répété. À cette fin, nous nous appuyons sur le modèle de Charnes et Granot, ainsi que sur l'analogie développée en section 2.2.1 entre la formation de coalitions stochastique répétée et les bandits manchots. Cela nous permet par la suite d'utiliser des stratégies d'exploration-exploitation dans la résolution de nos jeux de coalitions, en l'intégrant dans un concept de solutions bien connu dans la formation de coalitions, le cœur. Cela nous permet également de proposer deux nouveaux concepts de solutions : le γ-cœur et le δ-cœur. Nous évaluons ces deux nouveaux concepts en les opposant à une décision aléatoire et au concept de l'ϵ-cœur (couplé à la stratégie ϵ-gloutonne à des fins d'exploration). De plus, nous expérimentons l'impact du taux d'ex-

ploration dans un compromis exploration-exploitation pour la formation de coalitions. Ces expériences sont réalisées dans un premier temps avec un modèle d'estimation des utilités fondé sur une connaissance *a priori*, puis grâce à un modèle de réseau de neurones inspiré de ceux étudiés dans la littérature à propos des bandits manchots (voir section 2.2.3.3).

3.1 Jeux de coalitions stochastiques répétés

Dans le but de répondre à notre problématique de formation de coalitions dans un contexte stochastique où les agents n'ont pas de connaissances *a priori*, nous proposons un modèle qui intègre d'une part de l'incertitude mais aussi une notion explicite de temporalité. En effet, la temporalité nous permet de représenter la mise à jour des croyances des agents sur la fonction caractéristique au fur et à mesure de leurs interactions. Afin d'aborder la problématique sans ajouter de difficultés supplémentaires, le modèle proposé est centralisé et fermé (voir section 1.1.2). En conséquence, le modèle de croyances des agents est commun à ces derniers et de taille fixe (c'est-à-dire qu'aucune coalition ne sera rajoutée ou supprimée des croyances).

3.1.1 Fonction caractéristique stochastique et temporalité

En ce qui concerne l'incertitude, notre modèle s'appuie sur celui proposé par Charnes et Granot [Charnes et Granot, 1973] que nous avons présenté en section 2.1.3. Nous faisons ici l'hypothèse que les variables aléatoires décrivant les utilités des coalitions suivent des lois normales. En ce qui concerne la temporalité, nous nous appuyons sur l'analogie entre formation de coalitions et les bandits manchots développée en section 2.2.1. Dans ce jeu de coalitions, nous incorporons un ensemble de pas de temps $\mathbb{T}$ durant lesquels les agents font des observations sur les utilités réelles produites par les coalitions, qu'ils utilisent pour construire une fonction caractéristique estimée $\hat{v}$. Les solutions proposées par les agents sont donc fondées sur cette fonction caractéristique estimée, tout comme le sont les concepts de solutions. Formellement, un jeu de coalitions stochastique répété est défini comme suit :

Définition 3.1 (Jeu de coalitions stochastique répété)**.** *Soit* $\mathcal{G} = \langle N, \mathbb{T}, v, \hat{v} \rangle$ *un* jeu de coalitions stochastique répété *(RSCG, pour* Repeated Stochastic Coalitional Game*) où :*

— $N = \{a_1 \dots a_n\}$ *est un ensemble d'agents,*

— $\mathbb{T} \subset \mathbb{N}^+$ *est un ensemble de pas de temps discrets et distincts,*

— $v : 2^N \to \mathcal{X}^{2^N}$ *est une fonction caractéristique qui associe à chaque coalition une variable aléatoire. Pour une coalition donnée* $C \subseteq 2^N$*, nous notons* $v(C) = \mathcal{X}^C$ *où* $\mathcal{X}^C \sim \mathcal{N}^C(\mu_C, \sigma_C)$*. Cette fonction caractéristique est inconnue des agents.*

— $\hat{v} : 2^N \times \mathbb{T} \to \mathbb{R}$ *est une fonction caractéristique qui associe à chaque coalition au pas de temps* $t \in \mathbb{T}$ *une estimation de l'espérance d'utilité. Pour une coalition donnée* $C \subseteq 2^N$*, nous notons* $\hat{v}(C, t)$ *cette utilité estimée au pas de temps* t*. Cette fonction caractéristique représente les croyances communes des agents à propos de* v*, la fonction caractéristique réelle, au pas de temps* $t \in \mathbb{T}$*.*

Ici, l'utilité d'une coalition est donnée par une variable aléatoire qui suit une loi normale. Ce modèle comprend un ensemble de pas de temps $\mathbb{T}$ où les agents doivent décider pour chacun d'une solution à partir de la fonction caractéristique estimée $\hat{v}$. Une fois cette solution décidée et les coalitions formées, les agents observent les utilités produites et peuvent ensuite estimer les distributions de probabilités sous-jacentes, afin de mettre à jour la fonction caractéristique estimée $\hat{v}$ et ainsi recommencer au pas de temps suivant. Lorsque les agents doivent raisonner dans un cadre stochastique, il semble naturel que ces derniers prennent une décision en fonction de leur espérance de gain qui, par définition, représente le gain moyen espéré. C'est pourquoi la fonction caractéristique estimée associe à chaque coalition une espérance uniquement. De plus, afin d'exploiter l'analogie avec les bandits manchots au maximum, les utilités de la fonction caractéristique réelle sont normalisées sur l'intervalle $[0, 1]$. Le modèle présenté ci-dessus s'affranchit donc bien des hypothèses de déterminisme et de connaissance *a priori*. Cependant, il nous faut définir comment les agents construisent la fonction caractéristique estimée.

3.1.2 Estimation de la fonction caractéristique

Comme mentionné ci-dessus, les agents ne connaissent pas la fonction caractéristique réelle, et donc par extension les espérances d'utilités des coalitions non plus. Dans un tel contexte d'incertitude, les agents peuvent néanmoins profiter de la répétition du jeu afin de tenter d'estimer cette fonction caractéristique, tel qu'il est fait dans les modèles présentés en section 2.1.4. Il nous faut donc définir dans un premier temps ce qu'est une observation dans le cadre de notre modèle, puis comment les agents agrègent ces observations.

Supposons qu'à un pas de temps donné, une solution a été trouvée par les agents et les coalitions formées, les utilités réelles produites par les coalitions sont le résultat d'un processus stochastique paramétré par la fonction caractéristique, c'est-à-dire par les

variables aléatoires qu'elle décrit pour chaque coalition. Le modèle proposé étant dans un cadre centralisé, nous faisons l'hypothèse que toutes les utilités produites sont observées par l'ensemble des agents. Notons par $X_{t'}^{C}$ l'observation de l'utilité produite par la coalition C à un pas de temps t'.

Définition 3.2 (Observations). *Soit* $\mathcal{O}_t = \{(C, t', X_{t'}^{C}) : C \subseteq 2^N, t' \in \mathbb{T}, t' < t\}$ *un ensemble d'observations au pas de temps* t *correspondant à l'ensemble des coalitions formées à chaque pas de temps avant* t *et leurs utilités réelles produites. Étant donnée une structure de coalitions formées* $\mathcal{CS}^t$ *au pas de temps* t*, l'ensemble des observations au pas de temps* $t+1$ *est défini comme suit :*

$$\mathcal{O}_{t+1} = \mathcal{O}_t \cup \{(C, t, X_t^C) : \forall C \in \mathcal{CS}^t\}$$

Nous pouvons souligner le fait que l'ensemble des observations porte sur les pas de temps précédent t en excluant ce dernier car, à ce pas de temps, les agents estiment la fonction caractéristique v à partir de ces observations plus anciennes. De plus, définissons l'ensemble des observations qui ont été faites concernant une coalition particulière sur l'ensemble des pas de temps précédents (et ce pour la même raison que précédemment).

Définition 3.3 (Utilités observées pour une coalition). *Soit* $\mathcal{O}_t(C)$ *l'ensemble des utilités observées pour la coalition* $C \subseteq 2^N$ *au pas de temps* $t \in \mathbb{T}$*, tel que :*

$$\mathcal{O}_t(C) = \{X_{t'}^{C'} |\ \forall \{C', t', X_{t'}^{C'}\} \in \mathcal{O}_t \ t.q.\ C' = C\}$$

Afin que les agents puissent raisonner dans l'incertain, ils doivent être en capacité d'estimer les utilités des différentes coalitions, et ce à partir des observations préalablement mentionnées. Pour cela, nous proposons deux méthodes d'estimation différentes : une fondée sur une connaissance *a priori*, et une autre fondée sur une inférence.

3.1.2.1 Estimation à partir d'une connaissance *a priori*

Commençons par la méthode d'estimation fondée sur une connaissance *a priori*. Pour rappel, il est fait l'hypothèse dans le modèle que les distributions utilisées sont des lois normales, nous faisons donc également l'hypothèse que les agents savent cela. À partir des observations, les agents peuvent donc tenter d'estimer ces lois normales, et stockent dans la fonction caractéristique estimée $\hat{v}$ les espérances μ de ces lois.

Définition 3.4 (Estimation de l'utilité pour les lois normales)**.** *Soient C et $t \in \mathbb{T}$ respectivement une coalition et un pas de temps quelconques. L'estimation $\hat{v}(C,t)$ de l'utilité de la coalition C (dont l'utilité est définie par une variable aléatoire suivant une loi normale) au pas de temps t est définie telle que :*

$$\hat{v}(C,t) = \frac{\sum\limits_{x \in \mathcal{O}_t(C)} x}{|\mathcal{O}_t(C)|} \text{ avec } |\mathcal{O}_t(C)| > 0$$

Si $|\mathcal{O}_t(C)| = 0$, alors $\hat{v}(C,0)$ est tirée aléatoirement uniformément dans l'intervalle $[0,1]$.

Exemple 19. *Soient C une coalition, et $\mathcal{O}_t(C) = \{0.2, 0.4, 0.3\}$ l'ensemble des observations sur C au pas de temps t. L'estimation des agents est donc $\hat{v}(C,t) = 0.3$.*

3.1.2.2 Estimation par inférence

Pour cette méthode, nous ne faisons pas d'hypothèse sur la forme des lois et nous nous appuyons sur les travaux présentés en section 2.2.3.3, dans lesquels des problèmes de bandits manchots sont résolus à l'aide de réseaux de neurones. Tout comme il peut y avoir des dépendances entre certains bras pouvant être appris par les réseaux, il peut y avoir des dépendances entres coalitions et nous pouvons utiliser un réseau de neurones pour estimer la fonction caractéristique.

Définition 3.5 (Estimation par inférence)**.** *Une* estimation par inférence *est réalisée comme suit : à chaque pas de temps t', les agents observent les utilités des coalitions, et le réseau est entraîné avec l'ensemble des paires $(C, X^C_{t'})$ pour toute coalition C formée à ce pas de temps. La manière dont le réseau produit précisément les estimations dépend de l'algorithme d'optimisation utilisé [Burkart et Huber, 2021].*

Nous donnons lors de l'évaluation de notre modèle en section 4.1.2.3 un exemple d'instanciation concrète de ce réseau. À $t = 0$, en l'absence d'information, les valeurs des coalitions estimées par cette méthode sont déterminées par l'initialisation aléatoire des poids du réseau.

3.1.3 Solutions pour un RSCG

Étant donné que la fonction caractéristique réelle est inconnue des agents, ces derniers doivent donc résoudre le problème de formation des coalitions en s'appuyant sur les

estimations présentées ci-dessus. À l'instar du choix séquentiel de bras dans le problème des bandits manchots, les agents doivent ici décider à chaque pas de temps d'une solution au jeu, malgré le fait qu'ils ne connaissent pas *a priori* la fonction caractéristique. Une solution à un RSCG est, comme dans un contexte déterministe, un couple constitué d'une structure de coalitions et d'un vecteur de gains. Toutefois, d'une manière similaire à ce que Ieong et Shoham ont fait dans leur modèle présenté en section 2.1.3 [Ieong et Shoham, 2008], nous faisons la distinction d'une solution *ex-ante* à un RSCG, c'est-à-dire construite à partir des utilités estimées avant l'observation des utilités réelles, et une solution *ex-post* à un RSCG, c'est-à-dire après la formation des coalitions proposées dans la structure de la solution *ex-ante* et l'observation des utilités de ces dernières. La solution *ex-ante* étant calculée sur la base de gains estimés pour les agents, ces derniers calculent un contrat, calculé à partir des espérances de leurs coalitions respectives, qui correspond à la part d'utilité que chaque agent va gagner dans sa coalition. Les gains de la solution *ex-post* sont donc, pour un agent, l'utilité réelle observée de sa coalition pondérée par la part d'utilité espérée du contrat de la solution *ex-ante*.

Définition 3.6 (Solution *ex-ante* à un RSCG)**.** *Une* solution ex-ante S^t_{ante} *au pas de temps* $t \in \mathbb{T}$ *à un RSCG* $\mathcal{G}$ *est un couple* $S^t_{ante} = (\mathcal{CS}^t, \vec{x}^t_{ante})$ *tel que :*

- $\mathcal{CS}^t$ *est une structure de coalitions (partition disjointe) de* N*,*
- $\vec{x}^t_{ante} = \{\hat{x}^t_1, \ldots, \hat{x}^t_n\}$ *est un* contrat *représentant les parts d'utilité que les agents s'accordent à recevoir dans leurs coalitions respectives à l'observation des utilités réelles. Ce contrat est construit à partir des utilités* estimées*, tel que* $0 \leq \hat{x}^t_i \leq 1$ *est la part d'utilité pour l'agent* a_i *dans sa coalition* $C_i \in \mathcal{CS}^t$*, calculée selon l'utilité estimée* $\hat{v}(C_i, t)$*. Pour une coalition quelconque* $C \in \mathcal{CS}^t$*,* $\sum_{a_j \in C} \hat{x}^t_j = 1$*.*

Si une solution *ex-ante* à un RSCG est acceptée par tous les agents, une nouvelle solution *ex-post* à ce même RSCG est alors construite.

Définition 3.7 (Solution *ex-post* à un RSCG)**.** *Une* solution ex-post S^t_{post} *au pas de temps* $t \in \mathbb{T}$ *à un RSCG* $\mathcal{G}$ *est un triplet* $S^t_{post} = (S^t_{ante}, \mathcal{CS}^t, \vec{x}^t_{post})$ *tel que :*

- $S^t_{ante} = (\mathcal{CS}^t, \vec{x}^t_{ante})$ *est une solution* ex-ante *au pas de temps* t *à* $\mathcal{G}$*,*
- $\vec{x}^t_{post} = \{x^t_1, \ldots, x^t_n\}$ *est un vecteur de gain tel que* $x^t_i \geq 0$ *est le gain de l'agent* a_i *calculé selon l'utilité réelle* $X^{C_i}_t$ *de la coalition* C_i *à laquelle il appartient dans la structure* $\mathcal{CS}^t$ *et le contrat* $\vec{x}^t_{ante}$ *de la solution* ex-ante S^t_{ante}*, tel que :*

$$x^t_i = X^{C_i}_t \times \hat{x}^t_i$$

Exemple 20. *Soit un RSCG* $\mathcal{G} = \langle N, \mathbb{T}, v, \hat{v} \rangle$ *tel que* $N = \{a_1, a_2, a_3\}$ *et* $\mathbb{T} = \{0, \ldots, 99\}$. *Soit* $t = 21$ *où les fonctions caractéristiques réelles et estimées (avec* $\hat{v}$ *estimée à partir d'une connaissance* a priori *– Définition 3.4) sont respectivement :*

$$v = \{(a_1) = \mathcal{N}(0.4, 0.1)\ ;\ (a_2) = \mathcal{N}(0.3, 0.4)\ ;\ (a_3) = \mathcal{N}(0.2, 0.1)$$
$$(a_1, a_2) = \mathcal{N}(0.7, 0.2)\ ;\ (a_1, a_3) = \mathcal{N}(0.6, 0.1)\ ;\ (a_2, a_3) = \mathcal{N}(1, 0.2)$$
$$(a_1, a_2, a_3) = \mathcal{N}(0.7, 0.3)\ ;\ (\emptyset) = \mathcal{N}(0, 0)\}$$

$$\hat{v}(t) = \{(a_1) = 0.35\ ;\ (a_2) = 0.275\ ;\ (a_3) = 0.25$$
$$(a_1, a_2) = 0.65\ ;\ (a_1, a_3) = 0.625\ ;\ (a_2, a_3) = 0.95$$
$$(a_1, a_2, a_3) = 0.72\ ;\ (\emptyset) = 0\}$$

Une solution ex-ante *pourrait être* $S^t_{ante} = ([\{a_1\}\{a_2, a_3\}], \{x_1 : 1, x_2 : 0.29, x_3 : 0.71\})$. *Supposons alors les observations* $(\{a_1\}, t, 0.38)$ *et* $(\{a_2, a_3\}, t, 0.93)$, *la solution* ex-post *correspondante serait* $S^t_{post} = (S^t_{ante}, \{x_1 = 0.38, x_2 = 0.27, x_3 = 0.66\})$.

La définition d'une solution à un jeu de coalitions stochastique répété étant désormais établie, nous pouvons désormais nous intéresser à la définition de ce qu'est une solution stable dans un tel cadre, et ainsi proposer des concepts de solutions. De manière analogue aux bandits manchots, nous proposons de fonder nos concepts de solutions sur un équilibre entre exploration et exploitation.

3.2 γ-cœur : un ϵ-cœur biaisé par l'exploration

Pour créer un tel mécanisme, nous pouvons nous appuyer sur l'analogie développée en section 2.2.1 entre les bandits manchots et la formation de coalitions. En effet, dans le problème des bandits manchots, les stratégies s'appuyent sur un équilibre exploration-exploitation afin de ne pas délaisser des bras potentiellement intéressants. Nous pouvons donc nous appuyer sur ce principe afin de former des structures de coalitions qui pourraient être intéressantes. Dans cette optique, nous proposons deux nouveaux concepts de solutions exploratoires adaptés de l'ϵ-cœur. Le premier, appelé γ-cœur, intègre l'intérêt à l'exploration au même niveau que le gain des agents lors de la recherche de la stabilité, tandis que le deuxième, appelé δ-cœur, l'intègre sous la forme d'une part d'utilité sacrifiable. Commençons par construire le premier.

3.2.1 Adaptation de la stratégie UCB aux coalitions

Nous proposons donc de s'appuyer sur l'analogie effectuée entre les bandits manchots et la formation de coalitions pour proposer un biais d'exploration, à la manière d'un biais UCB dans les bandits manchots [Agrawal, 1995], représentant un intérêt à former la coalition pour obtenir plus d'informations sur son utilité réelle. Adaptons dans un premier temps le terme d'exploration d'UCB aux coalitions. Nous appelons ce biais, biais d'exploration coalitionnel.

Définition 3.8 (Biais d'exploration coalitionnel)**.** *Soit $\gamma_{coal}(C,t)$ un biais décrivant l'intérêt que peuvent avoir des agents à former une coalition $C \subseteq 2^N$ afin d'obtenir une estimation plus précise de son utilité aux pas de temps suivants. Ce biais est défini comme suit :*

$$\gamma_{coal}(C,t) = \sqrt{\frac{2.\log(|\mathcal{O}_t|+1)}{|\mathcal{O}_t(C)|+1}}$$

Cependant, lorsque les agents cherchent à déterminer si une solution est stable ou non, ils sont souvent amenés à comparer des structures différentes, et donc des coalitions composées de membres différents. Il nous faut donc considérer un biais sur des structures de coalitions. Cela peut être fait de manière analogue à la comparaison des gains dans le concept de solutions du cœur (voir Définition 1.13). Dans ce dernier, afin de trouver une structure de coalitions stable, les agents comparent leurs gains respectifs dans une solution proposée (donc des gains provenant de l'utilité de plusieurs coalitions) à l'utilité d'une coalition qu'ils pourraient créer en déviant de la solution (donc l'utilité d'une seule coalition). En appliquant le même principe, nous pouvons proposer un biais d'exploration *structurel*, c'est-à-dire qui prend en compte une solution proposée et agrège les biais d'exploration pour des agents n'étant pas nécessairement dans la même coalition. Cependant, nous pouvons nous interroger sur la nature des biais d'exploration agrégés. Prenons un exemple.

Exemple 21. *Soient S_1^t et S_2^t les solutions* ex-ante *S_{ante}^t au RSCG $\mathcal{G}$ (voir exemple 20 page 67) suivantes :*

$$S_1^t = ([\{a_1, a_2\}; \{a_3\}], \{x_1 : 0.55;\ x_2 : 0.45;\ x_3 : 1\}$$

$$S_2^t = ([\{a_1\}\{a_2, a_3\}], \{x_1 : 1, x_2 : 0.29, x_3 : 0.71\})$$

Si nous comparons les biais coalitionnels des coalitions de S_1^t aux biais coalitionnels des

coalitions de S_2^t, nous pouvons voir que ces coalitions ne sont pas identiques, et peuvent varier en taille et en nombre. Par exemple, si les agents a_1 et a_2 souhaitent évaluer l'intérêt à l'exploration qu'ils auraient en déviant de S_1^t à S_2^t, alors l'intérêt de a_3 sera également pris en compte, car il appartient à la nouvelle coalition de a_2 dans S_2^t. Toutefois, lors de cette comparaison, c'est bien uniquement l'intérêt des deux premiers agents qui nous intéresse. Il nous faut donc avoir une notion d'exploration individuelle.

Il semble donc pertinent de définir une part individuelle d'exploration dépendant d'une solution proposée.

Définition 3.9 (Part individuelle d'exploration)**.** *La* part individuelle d'exploration *d'un agent a_i dans une solution donnée $\mathcal{S}^t$ à un pas de temps t correspond à une part équitable du biais d'exploration de la coalition C_i à laquelle il appartient dans $\mathcal{CS}^t$, telle que :*

$$\gamma^i_{indiv}(\mathcal{CS}^t) = \frac{\gamma_{coal}(C_i, t)}{|C_i|}$$

Le biais d'exploration structurel pour une solution donnée agrège donc ces parts individuelles d'exploration afin de pouvoir comparer le biais d'exploration coalitionnel d'une coalition ciblée à l'intérêt exploratoire qu'ont les agents de cette coalition dans la solution donnée, qu'importe les coalitions dans lesquelles ils se trouvent.

Définition 3.10 (Biais d'exploration structurel)**.** *Soit C_i la coalition à laquelle appartient l'agent a_i dans une structure de coalition $\mathcal{CS}^t$. Le biais d'exploration* structurel *est défini comme suit :*

$$\gamma^{\mathcal{CS}^t}_{struct}(C, t) = \sum_{a_i \in C} \gamma^i_{indiv}(\mathcal{CS}^t)$$

Ce biais d'exploration structurel peut donc être intégré à un concept de solutions existant afin de prendre en compte une notion d'exploration.

Exemple 22. *Soient $\mathcal{G}$ le RSCG et S^t_{ante} la solution* ex-ante *définis dans l'exemple 20 (page 67) :*

$$S^t_{ante} = \left([\{a_1\}\{a_2, a_3\}], \{x_1 : 1, x_2 : 0.29, x_3 : 0.71\}\right)$$

Les biais d'exploration coalitionnels, les parts individuelles et les biais d'explorations structurels (par rapport à S^t_{ante}) selon le nombre d'observations pour chaque coalition (avec un nombre total d'observations de structures de coalitions de 21) sont indiqués dans le tableau 3.1. La part individuelle d'exploration permet donc de prendre en compte uniquement

C	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1, a_2, a_3\}$
$\lvert\mathcal{O}_t(C)\rvert$	12	13	17	6	2	1	1
$\gamma_{\text{coal}}(C,t)$	0.69	0.66	0.59	0.94	1.44	1.76	1.76
$\gamma_{\text{indiv}}(C,t)$	0.69	0.66	0.59	0.47	0.72	0.88	0.587
$\gamma^{\mathcal{CS}^t}_{\text{struct}}(C,t)$	0.69	0.88	0.88	1.57	1.57	1.76	2.45

TABLE 3.1 – Exemples de biais d'exploration coalitionnel et part individuelle

l'intérêt des agents concernés. Pour la solution S^t_{ante}*, si l'agent* a_1 *souhaite dévier et former sa coalition singleton, cela ne serait pas admis en comparant les biais d'exploration coalitionnels, car* $\gamma_{coal}(\{a_1\}, t) = 0.69$ *et* $\gamma_{coal}(\{a_1, a_2\}, t) = 0.94$*, ce qui prend en compte l'intérêt de* a_2*. Toutefois, si* a_1 *et* a_2 *formaient tous les deux leurs coalitions singleton (et donc la structure deviendrait* $[\{a_1\}\{a_2\}\{a_3\}]$*), ils auraient un meilleur intérêt exploratoire. En comparant la coalition souhaitée avec la part individuelle d'exploration,* a_1 *peut former sa coalition singleton, car* $\gamma_{coal}(\{a_1\}, t) = 0.69$ *tandis que* $\gamma_{indiv}(\{a_1, a_2\}, t) = 0.47$*. Cependant, lorsque nous raisonnons avec plus d'un agent, il est nécessaire d'agréger les parts individuelles avec le biais d'exploration structurel. Par exemple, dans la solution* S^t_{ante}*, le biais d'exploration structurel de la coalition* $\{a_1, a_2\}$ *est de* 1.57*, ce qui est supérieur au biais d'exploration coalitionnel de cette coalition, mais aussi au biais d'exploration structurel dans la structure* $[\{a_1\}\{a_2\}\{a_3\}]$*.*

3.2.2 Stabilité au sens du γ-cœur

Nous proposons d'adapter le concept de solution du ϵ-cœur (voir Définition 1.14) en considérant que la valeur d'une coalition, c'est-à-dire son intérêt à être formée à un pas de temps donné, dépend de deux éléments : une estimation de son utilité dont découle directement les gains des agents et son biais d'exploration coalitionnel.

Ainsi, une solution *ex-ante* d'un RSCG est *stable* au sens du *γ-cœur* si, et seulement si, il n'existe pas de coalitions qu'un groupe d'agents pourrait former, dont l'estimation de l'utilité plus le biais d'exploration coalitionnel soit supérieur à la somme des gains individuel des agents estimés dans cette solution plus le biais d'exploration structurel. De cette manière, le γ-cœur considère d'un côté ce que les agents gagneraient en termes de gains et d'intérêt à explorer dans une solution donnée, et de l'autre, l'utilité espérée et l'intérêt à explorer pour une coalition donnée, qui peut potentiellement faire dévier les agents. Formellement, cela est défini comme suit :

Définition 3.11 (γ-cœur)**.** *Une solution $S^t = (\mathcal{CS}^t, \vec{x}^t)$ d'un RSCG appartient au γ-cœur si, et seulement si :*

$$\forall C \in N, x^t(C) + \gamma^{\mathcal{CS}^t}_{struct}(C,t) \geq \hat{v}(C,t) - \epsilon + \gamma_{coal}(C,t)$$

tel que :

$$x^t(C) = \sum_{a_i \in C} \hat{x}^t_i \times \hat{v}(C_i, t)$$

où C_i est la coalition de l'agent a_i dans la solution S^t.

L'ajout de ces biais peut donc rendre stable dans le γ-cœur des structures de coalitions qui ne seraient pas stables dans un ϵ-cœur classique, et inversement. Il n'y a donc pas d'inclusion entre ces deux concepts. Prenons deux solutions en exemple.

Exemple 23. *Soient le jeu $\mathcal{G}$ défini dans l'exemple 20 (page 67) et les biais pour ce jeu calculés dans l'exemple 22 (page 69). Supposons une solution S^t_1 telle que :*

$$S^t_1 = ([\{a_1\}; \{a_2, a_3\}], \{x_1 : 1;\ x_2 : 0.29;\ x_3 = 0.71\}$$

Supposons une coalition $C = \{a_1, a_2\}$ telle que $\hat{v}(C,t) = 0.65$. La solution S^t_1 n'appartient pas au 0-cœur car :

$$x^t(C) = (\hat{x}^t_1 \times \gamma_{coal}(\{a_1\})) + (\hat{x}^t_2 \times \gamma_{coal}(\{a_2, a_3\}))$$
$$x^t(C) = (1 \times 0.35) + (0.29 \times 0.95) = 0.625$$

Nous avons donc $x^t(C) < \hat{v}(\{a_1, a_2\}, t)$. La coalition C empêche donc la solution S^t_1 d'appartenir au 0-cœur. Si nous calculons et ajoutons les biais coalitionnel et structurel pour vérifier si cette coalition C est dans le γ-cœur, nous avons :

$$\gamma^{\mathcal{CS}^t}_{struct}(C,t) = 0.69 + 0.88 = 1.57$$
$$\gamma_{coal}(C,t) = 0.94$$

Nous avons donc $x^t(C) + \gamma^{\mathcal{CS}^t}_{struct}(C,t) = 2.195 > \hat{v}(C,t) + \gamma_{coal}(C,t) - 0.4 = 1.59$. La coalition C qui empêchait la solution S^t_1 d'appartenir au 0-cœur ne l'empêche pas d'appartenir au γ-cœur. Aucune autre coalition n'empêche la stabilité de S^t_1 dans ce concept. Le γ-cœur ne contient donc pas l'ϵ-cœur.

Considérons maintenant S_2^t *la solution* ex-ante *suivante qui appartient au* $0,4$*-cœur :*

$$S_2^t = ([\{a_1, a_2\}; \{a_3\}], \{x_1 : 0.54;\ x_2 : 0.46;\ x_3 : 1\})$$

Soit la coalition $C = \{a_1, a_3\}$*. Attention, les biais structurels sont différents de ceux de l'exemple 22 pour cette solution.*

$$\hat{v}(C,t) = 0.625\ ;\ \gamma_{coal}(C,t) = 1.44$$
$$x^t(C) = (0.54 \times 0.65) + (1 \times 0.25) = 0.6\ ;\ \gamma_{struct}^{\mathcal{CS}^t}(C,t) = 0.47 + 0.59 = 1.06$$

Nous avons donc : $x^t(C) + \gamma_{struct}^{\mathcal{CS}^t}(C,t) = 1.66 < \hat{v}(C,t) + \gamma_{coal}(C,t) - 0.4 = 1.665$*. La solution* S_2^t *n'appartient donc pas au* γ*-cœur, qui n'est donc pas contenu dans l'*ϵ*-cœur.*

Nous pouvons également remarquer que, lorsque les agents ne possèdent aucune information, les biais sont égaux à 0 (car $log(1) = 0$) et le γ-cœur revient alors à un ϵ-cœur classique. Cependant, ce n'est pas le cas si toutes les coalitions ont été formées un nombre égal de fois non nul (i.e. les biais d'exploration coalitionnels sont donc égaux) car les parts individuelles d'exploration seront plus petites dans les grandes coalitions.

Nous pouvons cependant souligner certains problèmes. Pour commencer, nous intégrons dans le concept de solutions des éléments qui ne sont pas de l'utilité mais qui sont additionnés à de l'utilité pour déterminer si une solution est stable. Bien que cela n'impacte pas directement le vecteur de gain (les gains du vecteur restent inchangés), cela peut en revanche avoir un impact sur l'acceptation de certain de ces vecteurs, et ce au delà de l'intérêt à explorer. Prenons un exemple.

Exemple 24. *Soient les deux solutions* ex-ante *au RSCG défini dans l'exemple 20 (page 67) suivantes :*

$$S_1^t = ([\{a_1\}; \{a_2, a_3\}], \{x_1 : 1;\ x_2 : 0.29;\ x_3 : 0.71\}$$

$$S_2^t = ([\{a_1\}; \{a_2, a_3\}], \{x_1 : 1;\ x_2 : 0.37;\ x_3 : 0.63\}$$

telles que S_1^t *et* S_2^t *sont respectivement non-stables et stables dans le* 0*-cœur. Pour* S_1^t*, cela vient du fait que* $\hat{v}(\{a_1, a_2\}, t) > x^t(C)$ *avec* $\hat{v}(\{a_1, a_2\}, t) = 0.65$ *et* $x^t(C) = 0.625$*. Cependant,* S_1^t *est stable dans le* γ*-cœur car* $\gamma_{struct}^{\mathcal{CS}^t} = 1.57$ *et* $\gamma_{coal}(C,t) = 0.94$*, donc* $x^t(C) + \gamma_{struct}^{\mathcal{CS}^t} > \hat{v}(\{a_1, a_2\}, t) + \gamma_{coal}(C,t)$*, et aucune autre coalition n'empêche la stabilité de* S_1^t*. La solution* S_2^t *est également stable dans le* γ*-cœur. La solution* S_1^t *est non-stable dans le* 0*-cœur car elle n'est pas équitable. Cependant, grâce à l'exploration, elle a été rendue stable.*

Toutefois, pour la même structure de coalitions (et donc le même gain d'information pour les agents), il existait un contrat plus intéressant. L'équilibre exploration-exploitation n'est donc pas optimal.

En additionant le biais d'exploration structurel aux gains des agents, ces gains et l'intérêt d'explorer ont le même poids dans la décision, ce qui peut revenir à faire accepter des solutions possiblement très mauvaises en termes de gains pour certains agents, et ce grâce à l'exploration, tandis qu'une même structure (rapportant le même intérêt à l'exploration donc) avec un meilleur contrat était possible, comme l'illustre l'exemple 24.

De plus, lors du calcul des parts individuelles d'exploration dans une coalition, nous divisons le biais d'exploration coalitionnel par la cardinalité de la coalition. Bien que cela fait sens pour la comparaison de l'intérêt à explorer, cela est une hypothèse forte, car rien ne nous garantit qu'un tel biais d'exploration est transférable entre les agents, et donc divisible. Nous pouvons par exemple argumenter que nous pouvons ne pas faire cette hypothèse, et donc ne pas faire de division, ce qui favoriserait dans un premier temps les coalitions de faible cardinalité, et que lorsqu'elles auront été suffisamment formées, les coalitions de cardinalité plus élevée auront un biais d'exploration coalitionnel suffisant pour désormais être choisies. Un dernier point qui peut être discuté est le fait que tous les agents d'une coalition n'ont peut-être pas un intérêt égal à la formation de cette dernière, et donc qu'une division équitable n'est pas pertinente.

3.3 δ-cœur : le sacrifice pour l'exploration

Afin de résoudre les problèmes soulignés sur notre précédent concept de solutions, nous choisissons une nouvelle approche, qui n'est pas sans rappeler la sémantique du ϵ dans le concept de l'ϵ-cœur (voir Définition 1.14) : le sacrifice d'utilité pour l'exploration.

Le principe est que les agents peuvent choisir de sacrifier une partie de leur utilité, et ce dans l'objectif de former une coalition afin de gagner de l'information sur celle-ci. Ce sacrifice permet notamment de rendre stable des solutions qui ne le seraient pas sinon. L'utilité qu'un agent accepte de ne pas gagner s'appuie sur deux éléments. Le premier élément est le *surplus*, qui est la différence entre le gain que l'agent obtiendrait si la solution est acceptée et le gain qu'il obtiendrait s'il formait sa coalition singleton. Le deuxième élément est un *facteur d'exploration*, qui détermine quelle fraction de son surplus l'agent accepte de ne pas gagner. Pour cette raison, la valeur maximale de ce facteur doit être de 1 car une valeur supérieure représenterait le fait que l'agent accepte de gagner moins que

ce qu'il gagnerait seul, ce qui n'est donc pas compatible avec la propriété de *rationalité* des agents.

3.3.1 Définition du surplus

Le *surplus* est donc le premier élément constituant le gain sacrifiable d'un agent. Pour calculer ce surplus, l'agent calcule simplement la différence entre son gain espéré dans une solution *ex-ante* donnée (calculé sur la base d'un contrat donc) et l'utilité estimée qu'il gagnerait s'il formait sa coalition singleton. Ainsi, cette caractérisation du surplus permet de déterminer si une solution est rationnelle pour tous les agents, tout en calculant le montant maximal d'utilité que les agents peuvent sacrifier sans rendre la solution irrationnelle.

Définition 3.12 (Surplus)**.** *Soit $\Omega^t(a_i, S^t)$ le surplus de l'agent a_i pour une solution donnée S^t au pas de temps $t \in \mathbb{T}$. Ce surplus est calculé comme suit :*

$$\Omega^t(a_i, S^t) = (\hat{x}_i^t \times \hat{v}(C_i, t)) - \hat{v}(\{a_i\}, t)$$

où C_i est la coalition de a_i dans S^t. Si le surplus est négatif, cela signifie que la solution donnée est irrationnelle *pour l'agent a_i, et donc qu'elle ne sera jamais stable.*

3.3.2 Nouveau biais d'exploration normalisé : le gain sacrifiable

Concernant le *facteur d'exploration*, il peut être défini de plusieurs façons. L'unique contrainte est qu'il doit représenter une part de surplus, donc que sa valeur maximale doit être de 1. Ce nouveau facteur d'exploration que nous proposons est fondé lui-aussi sur un biais d'exploration coalitionnel présenté en section 3.2.1. Cependant, ce biais d'exploration n'étant pas naturellement limité à 1 (un exemple avec une coalition observée une seule fois, pour un nombre total de 7 observations, donne une valeur de 1.44 environ), nous le normalisons pour en faire un facteur.

Définition 3.13 (Facteur d'exploration)**.** *Soit $\gamma_{norm}(C, t)$ un facteur d'exploration, c'est-à-dire un biais d'exploration coalitionnel après normalisation.*

$$\gamma_{norm}(C, t) = \frac{\gamma_{coal}(C, t)}{\max\limits_{C' \subseteq 2^N} \gamma_{coal}(C', t)}$$

Exemple 25. *Reprenons les biais d'exploration coalitionnels montrés dans l'exemple 22 (page 69). Le facteur d'exploration pour chacun de ces biais est montré dans le tableau 3.2.*

C	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_2, a_3\}$	$\{a_1, a_2, a_3\}$
$\lvert\mathcal{O}_t(C)\rvert$	12	13	17	6	2	1	1
$\gamma_{\text{coal}}(C,t)$	0.69	0.66	0.59	0.94	1.44	1.76	1.76
$\gamma_{\text{norm}}(C,t)$	0.392	0.375	0.34	0.53	0.82	1	1

TABLE 3.2 – Exemples de biais normalisés

Ces biais normalisés étant calculés par rapport au biais d'exploration coalitionnel maximal dans le jeu, nous n'avons pas besoin d'un biais d'exploration structurel normalisé. En effet, tous les agents se réfèrent au même biais d'exploration maximal pour calculer leur intérêt à explorer.

Ce biais permet alors d'exprimer quelle part de leur surplus les agents sont prêts à sacrifier afin de gagner de l'information sur une coalition donnée. Nous pouvons remarquer que, si ce biais vaut 0, cela signifie que la coalition ne possède aucun intérêt pour l'exploration, et sa formation ne dépend donc que de son intérêt à l'exploitation. Il est important de noter que dans le cas où aucune coalition n'a été formée pour le moment, c'est-à-dire que les agents n'ont aucune observation, tous les facteurs d'exploration sont de 1. Un tel cas signifie que les agents sont prêts à abandonner tout leur surplus afin d'explorer n'importe quelle coalition. Une fois que tous les facteurs d'exploration ont été calculés, les agents peuvent les utiliser pour déterminer quelle part de surplus ils acceptent de ne pas gagner pour chaque coalition.

Définition 3.14 (Gain sacrifiable)**.** *Le* gain sacrifiable *pour un agent a_i et une solution donnée S^t au pas de temps t est donné par :*

$$\delta^t(a_i, S^t) = \Omega^t(a_i, S^t) \times \gamma_{norm}(C, t)$$

où C est la coalition de a_i dans S^t.

L'exemple suivant illustre les surplus et gains sacrifiables pour deux solutions différentes, avec un exemple de calcul pour un agent.

Exemple 26. *Soient S_1^t et S_2^t deux solutions* ex-ante *au RSCG $\mathcal{G}$. Notons que S_2^t est la même solution S_2^t telle que définie dans l'exemple 20 (page 67).*

$$S_1^t = ([\{a_1, a_2\}; \{a_3\}], \{x_1 : 0.55;\ x_2 : 0.45;\ x_3 : 1\}$$

$$S_2^t = ([\{a_1\}; \{a_2, a_3\}], \{x_1 : 1;\ x_2 : 0.37;\ x_3 : 0.63\}$$

Supposons les facteurs d'exploration calculés dans l'exemple 25 (page 75). Les surplus et gains sacrifiables pour les agents dans la solution S_1^t sont :

a_1 *: $\Omega^t(a_1, S_1^t) = (x_1^t \times \hat{v}(C_1, t)) - \hat{v}(\{a_1\}) = (0.55 \times 0.65) - 0.35 = 0.01$*
où $\delta^t(a_1, S_1^t) = \Omega^t(a_1, S_1^t) \times \gamma_{norm}(C_1, t) = 0.005$

a_2 *: $\Omega^t(a_2, S_1^t) = 0.015$ où $\delta^t(a_2, S_1^t) = 0.008$*

a_3 *: $\Omega^t(a_3, S_1^t) = \delta^t(a_3, S_1^t) = 0$*

Le surplus et le gain sacrifiable de l'agent a_3 ont la même valeur car a_3 est dans sa coalition singleton et il possède donc un surplus nul. Pour la solution S_2^t, les surplus et gains sacrifiables sont :

a_1 *: $\Omega^t(a_1, S_2^t) = \delta^t(a_1, S_2^t) = 0$*

a_2 *: $\Omega^t(a_2, S_2^t) = \delta^t(a_2, S_2^t) = 0.075$*

a_3 *: $\Omega^t(a_3, S_2^t) = \delta^t(a_3, S_2^t) = 0.35$*

Les valeurs sont identiques dans S_2^t car a_1 est dans sa coalition singleton, et la coalition $\{a_1, a_2\}$ a un facteur d'exploration de 1.

Ce gain sacrifiable peut donc désormais être utilisé dans un concept de solutions afin de rendre certaines solutions stables par rapport à un ϵ-cœur classique.

3.3.3 Stabilité au sens du δ-*cœur*

En adaptant à nouveau le concept de solutions du cœur (voir Définition 1.14), nous proposons un nouveau concept de solutions dans lequel former une structure de coalitions *stable* signifie qu'il n'existe aucune coalition qui ne fait pas partie de la structure pour laquelle sa valeur estimée, moins la somme des gains que les agents acceptent de ne pas gagner afin de former la structure, est supérieure à la somme des gains que les agents estiment qu'ils vont gagner dans la structure (calculé grâce au contrat de la solution *ex-ante*). Nous appelons ce concept de solutions le δ-*cœur*.

Définition 3.15 (δ-core)**.** *Une solution* $S^t = (\mathcal{CS}^t, \vec{x}^t)$ *à un RSCG est* stable *si, et seulement si :*

$$\forall C \in N, x^t(C) \geq \hat{v}(C,t) - \epsilon - \Delta^t(C)$$

avec :

$$x^t(C) = \sum_{a_i \in C} \hat{x}_i^t \times \hat{v}(C_i, t)$$

où C_i *est la coalition de l'agent* a_i *dans la solution* S^t*, et :*

$$\Delta^t(C) = \sum_{a_i \in C} \delta^t(a_i, S^t)$$

Avec ce nouveau concept de solutions, nous n'intégrons plus d'éléments extérieurs à l'utilité, mais seulement la somme des utilités renoncées par les agents.

Exemple 27. *Soient les deux solutions* ex-ante *au RSCG défini dans l'exemple 20 (page 67) :*

$$S_1^t = ([\{a_1\}; \{a_2, a_3\}], \{x_1 : 1;\ x_2 : 0.29;\ x_3 : 0.71\}$$

$$S_2^t = ([\{a_1\}; \{a_2, a_3\}], \{x_1 : 1;\ x_2 : 0.37;\ x_3 : 0.63\}$$

Considérons les surplus et gains sacrifiables calculés dans l'exemple 26 (page 75). La solution S_2^t *est stable dans le δ-cœur, tandis que* S_1^t *ne l'est pas. Prenons un exemple de calcul pour cette dernière. Considérons la coalition* $C = \{a_1, a_2\}$ *:*

$$\Delta^t(C) = (\Omega^t(a_1, S_1^t) \times \gamma_{norm}(C_1, t)) + (\Omega^t(a_2, S_1^t) \times \gamma_{norm}(C_2, t)) = (0 \times 0.392) + (0 \times 1) = 0$$
$$\hat{v}(C,t) - \Delta^t(C) = 0.65 - 0 = 0.65$$
$$x^t(C) = (\hat{x}_1^t \times \hat{v}(C_1, t)) + (\hat{x}_2^t \times \hat{v}(C_2, t)) = (1 \times 0.35) + (0.29 \times 0.95) = 0.625$$

Ainsi, $\hat{v}(C,t) - \Delta^t(C) > x^t(C)$*. La solution* S_1^t *est donc instable en raison de la coalition* C*. Contrairement au γ-cœur, la solution* S_1^t *n'est pas stable dans le δ-cœur (voir l'exemple 23 page 71). Dans cette solution* S_1^t*, il est évident que* a_2 *est lésé, comparé à la solution* S_2^t *qui est plus juste. Le δ-cœur présente donc ici un équilibre exploration-exploitation plus satisfaisant.*

Contrairement au γ-cœur, il n'est pas fait l'hypothèse que l'intérêt à explorer puisse être équivalent à de l'utilité transférable, en effet, ici cet intérêt ne sert qu'à pondérer le gain qui peut être sacrifié pour l'exploration. De plus, le fait de pondérer individuellement

les gains des agents permet aux agents d'une même coalition de choisir de sacrifier une part différente de leur gain, ce qui a pour conséquence de rendre l'intérêt exploratoire pour une coalition différent entre les agents de cette dernière. Cependant, un problème peut être identifié dans ce nouveau concept de solutions. En effet, la construction du facteur d'exploration normalisé est telle qu'il existe *toujours* une coalition pour laquelle ce facteur est maximal, c'est-à-dire ayant la valeur 1. Cela signifie que à chaque pas de temps du problème, les agents peuvent sacrifier tout leur gain pour l'exploration de la coalition la moins formée (ou de plusieurs coalitions si tel est le cas), et ce à horizon infini, tandis que le biais d'exploration UCB des bandits manchots sur lequel nous nous fondons est censé décroître jusqu'à devenir nul lorsque le nombre d'observations total augmente. Cette modification de comportement peut donc provoquer un déséquilibre entre l'exploration et l'exploitation.

Chapitre 4

Comparaison des concepts de solutions fondés sur l'exploration

Sommaire

4.1 Expérimentations

Afin de déterminer si les concepts de solutions γ-cœur et δ-cœur que nous proposons sont intéressants pour résoudre des jeux de coalitions stochastiques répétés, nous procédons à une analyse empirique. Nous comparons notamment nos concepts de solutions au concept de solutions classique ϵ-cœur.

Pour cela, nous générons des jeux aléatoires que les agents jouent de manière répétée et observons l'évolution de l'optimalité des solutions choisies, et ce en s'appuyant sur trois métriques : le regret instantané, le regret cumulé et l'erreur moyenne absolue.

Bien que l'hypothèse est faite dans le modèle que les utilités des coalitions sont des variables aléatoires suivant des lois normales, il n'en est rien de la façon de les construire.

Étant donné que la structure de la fonction caractéristique peut avoir un impact significatif sur la résolution du jeu (voir section 1.3.2.1), nous définissons ci-après un protocole expérimental utilisant différentes structures possibles. De plus, l'apprentissage de la fonction caractéristique étant un aspect important des RSCG, les expériences menées le sont avec nos deux modèles d'estimations : l'approche fondée sur une connaissance *a priori* et l'approche fondée sur l'inférence, présentées au chapitre précédent.

4.1.1 Classes de fonctions caractéristiques

Afin d'évaluer nos concepts de solutions, nous construisons des jeux de coalitions stochastiques répétés ayant des fonctions caractéristiques structurées de différentes manières, regroupées en *classes* de fonctions caractéristiques. Ainsi, pour chaque jeu construit, la classe de fonctions caractéristiques utilisée détermine le modèle selon lequel seront déterminés les paramètres pour construire les variables aléatoires correspondant à chaque coalition.

La première classe de fonctions caractéristiques que nous utilisons regroupe les fonctions caractéristiques construites suivant le modèle NDCS (Normally Distributed Coalition Structures), proposé par Rahwan *et al.* [Rahwan *et al.*, 2009]. Cette classe a initialement été proposée dans le cadre de l'évaluation des différents algorithmes de génération de structures de coalitions (voir section 1.3.3.1). Lorsque que Rahwan *et al.* ont proposé la classe NDCS, ils l'ont comparée à deux autres classes, *Normal* et *Uniform*, proposées par Larson et Sandholm [Larson et Sandholm, 2000].

Ces différentes classes de fonctions caractéristiques ont toutefois été proposées dans un cadre déterministe de formation de coalitions, c'est-à-dire où les utilités des coalitions sont des réels. Dans notre cadre stochastique, le modèle construit alors les espérances de valeur pour chaque coalition. Autrement dit, pour une coalition C, la valeur créée est le paramètre μ_C de la loi normale $\mathcal{N}^C(\mu_C, \sigma_C)$, que la valeur aléatoire $\mathcal{X}^C$ suit.

4.1.1.1 NDCS

Le principe de ce modèle est de structurer la fonction caractéristique (ici dans un cadre déterministe) de telle façon que plus une coalition est grande, plus sa valeur espérée est élevée. Cependant, afin d'éviter de créer une structure monotone ou superadditive (voir Section 1.3.2.1), plus une coalition est grande, plus sa valeur espérée peut être soumise à une variance importante. Formellement, cela se traduit par une espérance et une variance

qui croissent avec la cardinalité de la coalition.

Définition 4.1 (Classe *NDCS*)**.** *Soit une coalition C, la valeur μ_C de sa loi normale $\mathcal{N}^C$ est déterminée par un tirage aléatoire de la variable aléatoire $\mathcal{X}$ suivant une loi normale telle que :*

$$\mu_C = \mathcal{X} \sim \mathcal{N}(|C|, \sqrt{|C|})$$

4.1.1.2 Normal

Ici, bien que l'utilité espérée d'une coalition dépende toujours de sa cardinalité, cette dernière est pondérée par le résultat d'un tirage d'une variable aléatoire suivant une loi normale. Comme chez Rahwan *et al.*, cette loi normale est paramétrée de manière identique pour toutes les coalitions. Il est cependant important de noter que ce tirage est répété et indépendant entre les différentes coalitions.

Définition 4.2 (Classe *Normal*)**.** *Soit une coalition C, la valeur μ_C de sa loi normale $\mathcal{N}^C$ est déterminée par sa cardinalité pondérée par un tirage aléatoire de la variable aléatoire $\mathcal{X}$ suivant une loi normale telle que :*

$$\mu_C = |C| \times \mathcal{X} \sim \mathcal{N}(1, 0.1)$$

4.1.1.3 Uniform

De la même manière que pour la classe *Normal*, l'utilité d'une coalition dépend de sa cardinalité pondérée par le résultat d'un tirage d'une variable aléatoire. Cependant dans ce cas-ci, le tirage suit une loi de probabilité uniforme. Tout comme pour la classe *Normal*, la loi est paramétrée de manière identique pour toutes les coalitions, et le tirage est répété et indépendant entre chacune de ces dernières.

Définition 4.3 (Classe *Uniform*)**.** *Soit une coalition C, la valeur μ_C de sa loi normale $\mathcal{N}^C$ est déterminée par sa cardinalité pondérée par un tirage aléatoire de la variable aléatoire $\mathcal{X}$ suivant une loi uniforme telle que :*

$$\mu_C = |C| \times \mathcal{X} \sim \mathcal{U}(0, 1)$$

4.1.1.4 Random

Enfin, nous utiliserons une dernière classe de modèle, à savoir *Random*. Dans cette classe, nous nous affranchissons de la taille des coalitions pour construire les utilités de

ces dernières, et ce afin d'évaluer l'impact sur la décision qu'ont les structures fondées sur la cardinalité des coalitions. Formellement, l'utilité de chaque coalition est déterminée par un tirage d'une variable aléatoire suivant une loi de probabilité uniforme. À nouveau, cette dernière est paramétrée de manière identique pour toutes les coalitions, et le tirage est répété et indépendant entre chacune de ces dernières.

Définition 4.4 (Classe *Random*)**.** *Soit une coalition C, la valeur μ_C de sa loi normale $\mathcal{N}^C$ est déterminée par un tirage aléatoire de la variable aléatoire $\mathcal{X}$ suivant une loi uniforme telle que :*

$$\mu_C = \mathcal{X} \sim \mathcal{U}(0, 1)$$

4.1.2 Déroulement des expérimentations

L'objectif étant d'évaluer nos concepts de solutions, il nous faut mettre en place un protocole d'expérimentation.

4.1.2.1 Paramètres des expérimentations

Nous générons dans un premier temps des jeux avec 5, 6 et 7 agents, pour lesquels il existe respectivement 52, 203 et 877 structures de coalitions possibles. Les fonctions caractéristiques de ces jeux décrivent des variables aléatoires suivant une loi normale dont le paramètre μ est construit selon une des quatre classes présentées ci-dessus. Concernant le paramètre σ de ces lois normales, nous faisons l'hypothèse que pour chaque coalition C, cette variance est égale à $\sigma_C = \mathcal{U}(0, \frac{\mu_C}{2})$. Étant donné que la variance maximale dépend de μ et que ce dernier tend à croître avec les coalitions de cardinalité plus élevée, alors plus une coalition contient d'agents, plus cette variance risque d'être grande, ce qui correspond intuitivement au fait que plus un groupe est grand, plus la collaboration au sein de celui-ci est soumise à aléas. Pour chacune des quatre classes de fonctions caractéristiques et chaque ensemble d'agents considérés, 1000 jeux différents (et donc des fonctions caractéristiques différentes, avec des variables aléatoires différentes) sont générés, durant lesquels les agents jouent pendant 100 pas de temps, c'est-à-dire $\mathbb{T} = [0, 99]$ pour chaque jeu.

4.1.2.2 Stratégie de comparaison

Nous mettons nos concepts de solutions en concurrence avec l'ϵ-cœur, dont ils sont une adaptation. Cependant, l'initialisation des croyances déterminera la solution trouvée par

l'ϵ-cœur, qui restera identique à chaque pas de temps suivant en l'absence de mise à jour des croyances, étant donné que ce concept est purement fondée sur l'exploitation. C'est pourquoi nous couplons ce concept à un stratégie d'exploration-exploitation ϵ-gloutonne, où l'exploitation est remplacée par le résultat de l'ϵ-cœur.

Cependant, l'exploration étant paramétrée par la valeur ϵ de la stratégie ϵ-glouton, nous pouvons nous interroger sur la meilleure valeur à utiliser pour paramétrer cette stratégie. C'est pourquoi nous présentons en section 4.2.1 des expérimentations concernant les performances de cette dernière lorsque sa valeur ϵ varie entre 0 et 1.

4.1.2.3 Modèle d'un réseau de neurones pour les RSCG

Afin de résoudre les jeux que nous avons construit, nous nous appuyons sur les travaux présentés en section 2.2.3.3, dans lesquels des problèmes de bandits manchots sont résolus à l'aide de réseaux de neurones.

Pour nous expérimentations, nous utilisons donc un réseau de neurones possédant deux couches cachées. Chaque couche est une couche dense avec une fonction d'activation ELU. La couche d'entrée représente les coalitions avec un neurone unique dédié à chaque agent : une valeur de 1 sur un tel neurone signifie la présence de l'agent dans la coalition, et une valeur de 0 son absence. La couche de sortie consiste en un unique neurone qui produit une valeur réelle. Cette valeur est donc l'utilité moyenne estimée pour la coalition entrée en paramètre du réseau, c'est-à-dire $\hat{v}(C, t)$ pour une coalition C au pas de temps t. Un tel réseau est capable d'apprendre des fonctions non-linéaires. Nous utilisons une descente de gradient stochastique avec estimations adaptatives des moments afin d'entraîner le réseau, tandis que la fonction d'erreur est l'erreur quadratique moyenne. Un schéma de ce réseau de neurones pour n agents est montré par la figure 4.1.

Il est important de souligner que ce réseau n'est pas entraîné hors ligne avant de jouer les RSCG (c'est-à-dire entraîné sur des données de test avant l'application au véritable problème) mais entraîné durant le problème à la manière de l'apprentissage par renforcement.

4.1.3 Métriques

Afin d'évaluer les performances de nos concepts de solutions, nous mesurons à la fois l'efficacité des décisions prises (c'est-à-dire l'optimalité de la solution stable trouvée par les agents) à travers le temps, et la précision de l'estimation que les agents font de la fonction

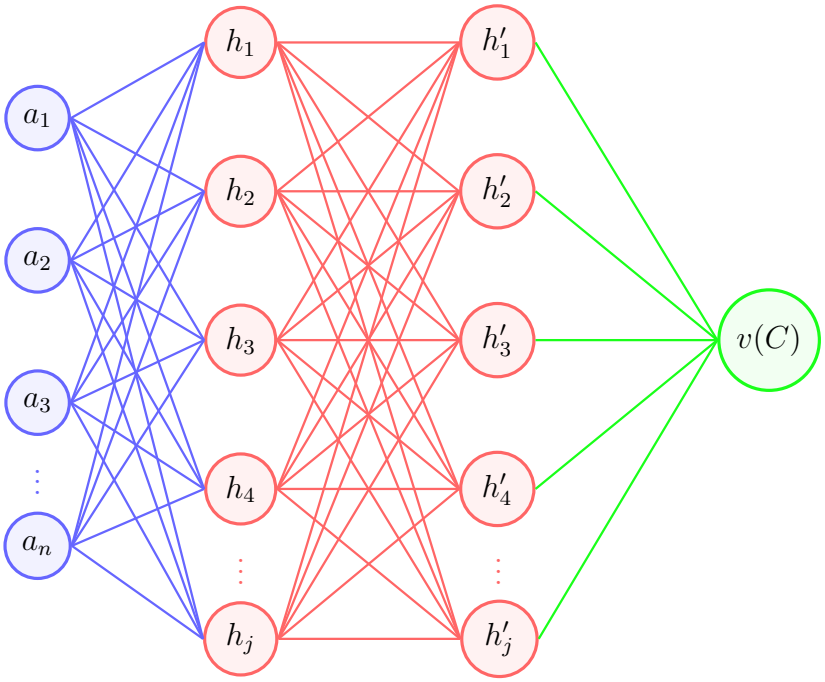


FIGURE 4.1 – Représentation du réseau de neurones utilisé

caractéristique réelle. Cela se traduit par des mesures de regret, mesure classique dans les bandits manchots, sachant que la minimisation du regret est équivalente à la maximisation des gains, et une mesure d'erreur pour l'apprentissage de la fonction caractéristique.

4.1.3.1 Regret instantané

La première mesure est le *regret instantané* qui est mesurée comme étant la différence entre le bien-être social maximal pour le jeu (voir la définition 1.19) et la somme des utilités réelles espérées des coalitions de la structure formée au pas de temps t. Formellement, le regret instantané est défini tel que :

Définition 4.5 (Regret instantané)**.** *Soit $S^* = (\mathcal{CS}^*, \vec{x}^*)$ la solution optimale au sens du bien-être social, le regret instantané au pas de temps t, noté R^t, est défini tel que :*

$$R^t = \sum_{C^* \in \mathcal{CS}^*} \mu_{C^*} - \sum_{C \in \mathcal{CS}^t} \mu_C$$

En raison de la stochasticité, le regret instantané peut osciller, parfois avec une grande amplitude, c'est pourquoi nous nous intéresserons principalement dans la suite à une

seconde mesure : le regret cumulé.

4.1.3.2 Regret cumulé

Le *regret cumulé* est une mesure permettant d'observer l'évolution du regret instantané au cours du temps. Cela est notamment utile afin d'étudier la convergence du regret, c'est-à-dire à partir de quel pas de temps les stratégies ont atteint leur équilibre exploration-exploitation et produisent donc un regret instantané de valeur constante. À un pas de temps t, il est défini comment étant la somme du regret instantané de chaque pas de temps précédent t, plus le regret instantané au pas de temps t. Formellement, le regret cumulé est défini tel que :

Définition 4.6 (Regret cumulé)**.** *Soit $S^* = (\mathcal{CS}^*, \vec{x}^*)$ la solution optimale au sens du bien-être social, le regret cumulé à un pas de temps t, noté R_c^t, est défini tel que :*

$$R_c^t = \sum_{t'=0}^{t} R^{t'}$$

4.1.3.3 Erreur moyenne absolue (MAE)

Enfin, afin d'évaluer l'estimation $\hat{v}$ que les agents font de la fonction caractéristique réelle v, nous proposons de définir une *distance entre fonctions caractéristiques*. Une telle distance peut être définie grâce à l'*erreur moyenne absolue* (MAE) sur les utilités estimées et réelles des coalitions. D'un point de vue pratique, plus la MAE est proche de 0, plus la fonction caractéristique estimée est précise. Formellement, la distance entre fonctions caractéristiques est définie telle que :

Définition 4.7 (Distance entre fonctions caractéristiques)**.** *Soient v et $\hat{v}$ deux fonctions caractéristiques, la distance D_{MAE}^t entre v et $\hat{v}$ au pas de temps t est définie telle que :*

$$D_{MAE}^t = \frac{\sum_{C \in 2^N} |\hat{v}(C) - v(C)|}{|2^N|}$$

Nous avons donc défini des mesures de regret des solutions trouvées par les agents, ainsi qu'une mesure de précision de l'estimation de la fonction caractéristique effectuée par ces derniers. Afin de rendre ces mesures commensurables entre les différentes expérimentations qui ne sont pas paramétrées par le même nombre d'agents, les regrets instantanés et cumulés sont divisés par le nombre d'agents.

4.2 Expérimentations : estimation sur une connaissance *a priori*

Dans un premier temps, les expérimentations sont effectuées avec la méthode d'estimation fondée sur une connaissance *a priori*. Étant donné qu'il semble évident que plus il y a d'exploration, plus nous nous rapprochons d'une simple stratégie aléatoire, les pas de la valeur ϵ sont de plus en plus espacés à mesure que celle-ci grandit. Pour ces expérimentations, nous générons également des jeux de 4 agents, paramétrés de la même manière que ceux de 5 à 7 agents.

4.2.1 Effets de l'exploration avec ϵ-glouton

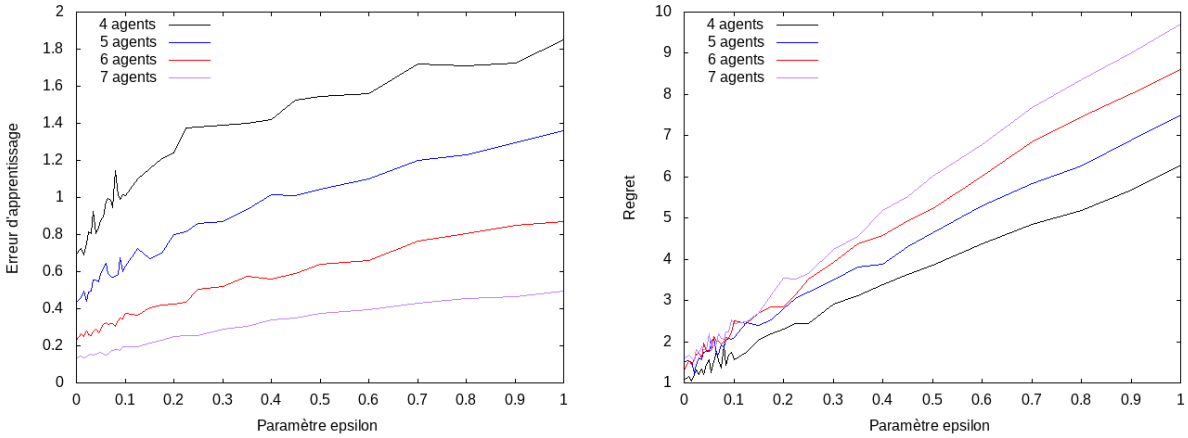
Ces expérimentations ont pour objectif d'évaluer les performances de la stratégie ϵ-glouton dont le terme d'exploitation est remplacé par le concept de solutions ϵ-cœur, lorsque nous faisons varier la valeur ϵ de la stratégie, en allant de 0 à 1. De la même manière que pour cette expérience dans un cadre sans inférence, les pas de la valeur ϵ sont de plus en plus espacés à mesure que celle-ci grandit et nous générons également des jeux de 4 agents, paramétrés de la même manière que ceux de 5 à 7 agents.

4.2.1.1 Graphiques

Les figures 4.2, 4.3, 4.4 et 4.5 présentent les résultats sur l'erreur d'apprentissage et le regret cumulé respectivement pour les classes *Normal*, *Uniform*, *NDCS*, et *Random*, et ce en fonction du nombre d'agents (allant de 4 à 7), avec en abscisse les valeurs d'ϵ pour la stratégie ϵ-glouton qui ont été testées.

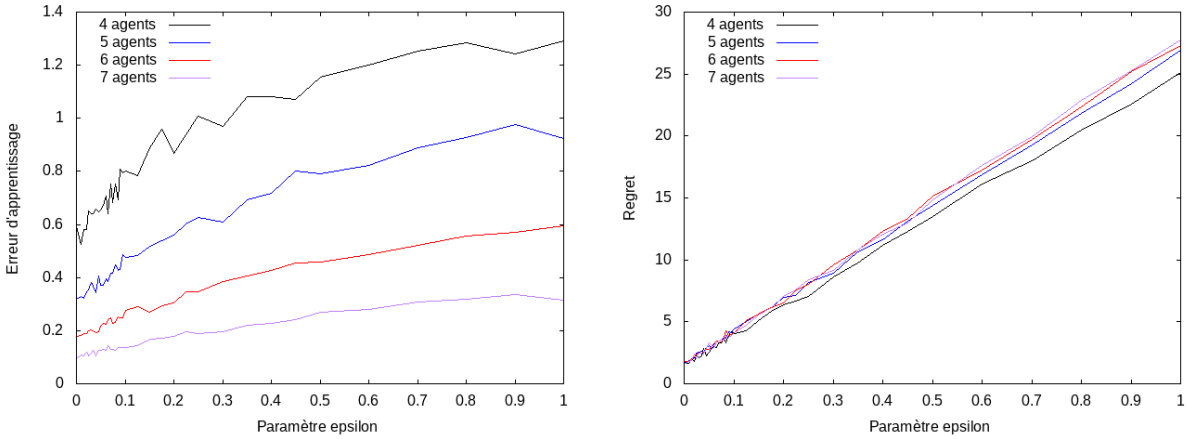
4.2.1.2 Analyse des résultats

Pour analyser les résultats, il convient de savoir que les nombreuses oscillations, lorsque ϵ est petit, sont dues au plus grand nombre de données sur l'intervalle $[0, 0.1]$. Nous pouvons dans un premier temps souligner le fait qu'indépendemment du nombre d'agents, et pour toutes les classes de fonctions caractéristiques utilisées, l'erreur d'apprentissage augmente lorsque ϵ croît, c'est-à-dire lorsque la stratégie dérive de très gloutonne à une exploration aléatoire. Cela paraît contre-intuitif mais cela peut s'expliquer par le fait que les agents initialisent une fonction caractéristique estimée de **même classe** que la fonction caractéristique réelle dont ils n'ont aucune connaissance. Cependant, le fait de connaître la classe et donc la structure de la fonction caractéristique permet aux agents de créer



(a) Erreur moyenne absolue finale (b) Regret cumulé

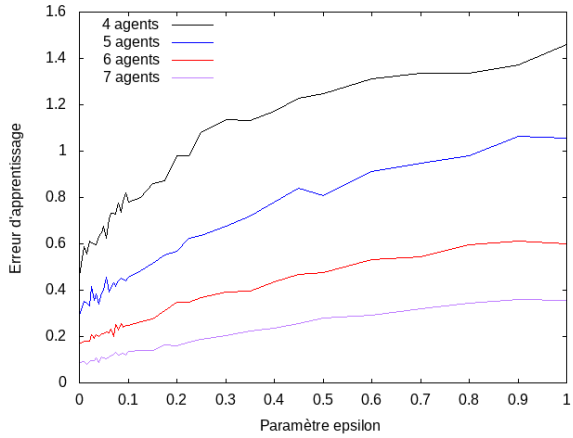
FIGURE 4.2 – Résultats pour la stratégie ϵ-gloutonne avec la classe Normal



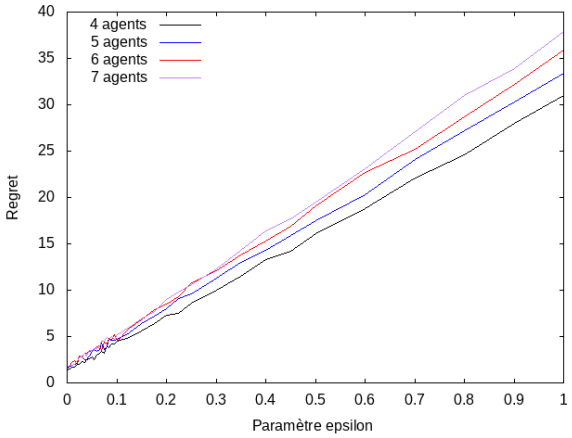
(a) Erreur moyenne absolue finale (b) Regret cumulé

FIGURE 4.3 – Résultats pour la stratégie ϵ-gloutonne avec la classe Uniform

une première estimation qui puisse être assez juste, ce qui est cohérent avec les valeurs d'erreur d'apprentissage que nous pouvons observer : si nous prenons l'ensemble des expérimentations, elles oscillent entre 0.1 et 1.9, ce qui en font des valeurs très basses. Lorsque nous prenons une stratégie davantage gloutonne, les agents vont former souvent les mêmes coalitions, ce qui fait que pour ces coalitions, de nombreuses observations permettent de réduire au maximum l'incertitude due à la variance de l'utilité des coalitions. En revanche, une stratégie exploratoire aura tendance à former de nombreuses coalitions peu de fois, et donc observer peu de valeurs pour chaque coalition, or, sachant la variance possible des utilités des coalitions, de mauvais tirages peuvent rendre l'estimation de la fonction carac-

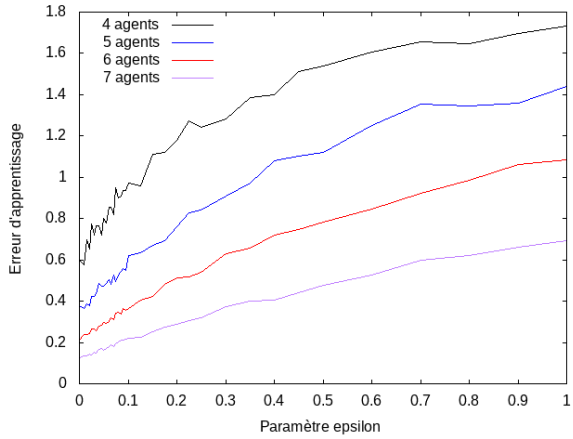


(a) Erreur moyenne absolue finale

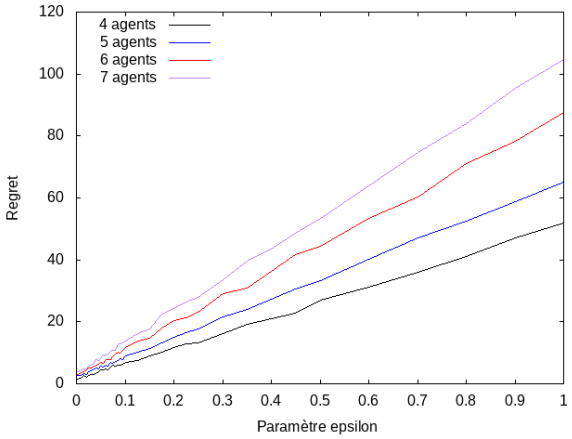


(b) Regret cumulé

FIGURE 4.4 – Résultats pour la stratégie ϵ-gloutonne avec la classe NDCS



(a) Erreur moyenne absolue finale



(b) Regret cumulé

FIGURE 4.5 – Résultats pour la stratégie ϵ-gloutonne avec la classe Random

téristique moins bonne si la première estimation était déjà correcte. Concernant le regret cumulé en revanche, celui-ci augmente très clairement lorsque ϵ croît. Pour les classes *Uniform* et *NDCS*, les valeurs absolues restent cependant très proches lorsque le nombre d'agents augmente, avec une légère tendance à l'augmentation (c'est-à-dire que plus il y a d'agents, plus le regret cumulé est élevé). La différence est bien plus importante pour les classes *Normal* et *Random*, avec toujours des valeurs absolues plus grandes lorsque le nombre d'agents augmente. Cela reste cependant intuitif, car plus il y a d'agents, plus il y a des pertes d'utilité possibles. Ainsi, bien que l'exploration avec une méthode d'estimation sur une connaissance *a priori* ne semble pas être intéressante que ce soit pour

l'apprentissage des fonctions caractéristiques, ou bien en termes de regret.

4.2.1.3 Conclusion intermédiaire

Contrairement à ce qui pourrait être attendu, davantage d'exploration ne permet pas un meilleur apprentissage de la fonction caractéristique réelle. Cela peut être expliqué par le fait que les agents intialisent une croyance déjà précise, et ce en raison de la connaissance *a priori* de la structure de la fonction caractéristique. En revanche, comme nous pouvions nous y attendre, lorsque la décision devient moins gloutonne, le regret augmente drastiquement.

4.2.2 δ-cœur contre γ-cœur, ϵ-glouton et une décision aléatoire

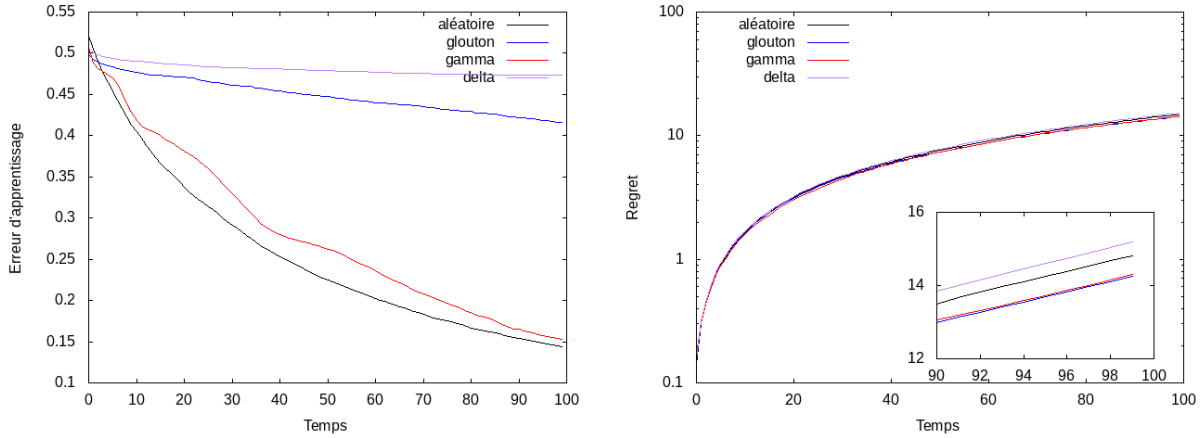
Nous souhaitons désormais mettre en concurrence les deux concepts de solutions que nous proposons, avec également avec une stratégie aléatoire ainsi que la stratégie fondée sur l'ϵ-cœur présentée en section 4.1.2.2. La méthode d'estimation utilisée est toujours celle fondée sur une connaissance *a priori*. Concernant la valeur ϵ de la stratégie ϵ-glouton dans les expérimentations suivantes, nous avons choisi de prendre une valeur arbitraire dans l'intervalle le plus intéressant à la vue des résultats, à savoir l'intervalle $[0, 0.1]$. Le choix s'est donc porté sur la valeur médiane de cet intervalle, tel que $\epsilon = 0.05$.

4.2.2.1 Graphiques

Les figures 4.6, 4.7, 4.8 et 4.9 représentent les expérimentations sur l'effet de l'exploration respectivement pour les classes *Normal*, *Uniform*, *NDCS* et *Random* pour 5 agents, les figures 4.10, 4.11, 4.28 et 4.13 celles pour les classes *Normal*, *Uniform*, *NDCS* et *Random* pour 6 agents, et enfin les figures 4.14, 4.15, 4.16 et 4.17 respectivement celles pour les classes *Normal*, *Uniform*, *NDCS* et *Random* pour 7 agents. Afin de permettre une meilleur lisibilité des courbes de regret lorsque celles-ci sont très proches, un agrandissement sur les 10 derniers pas de temps est fourni en supplément pour les classes *Normal*, *Uniform* et *NDCS*. Pour la classe *Random*, les courbes sont distinguables les unes des autres.

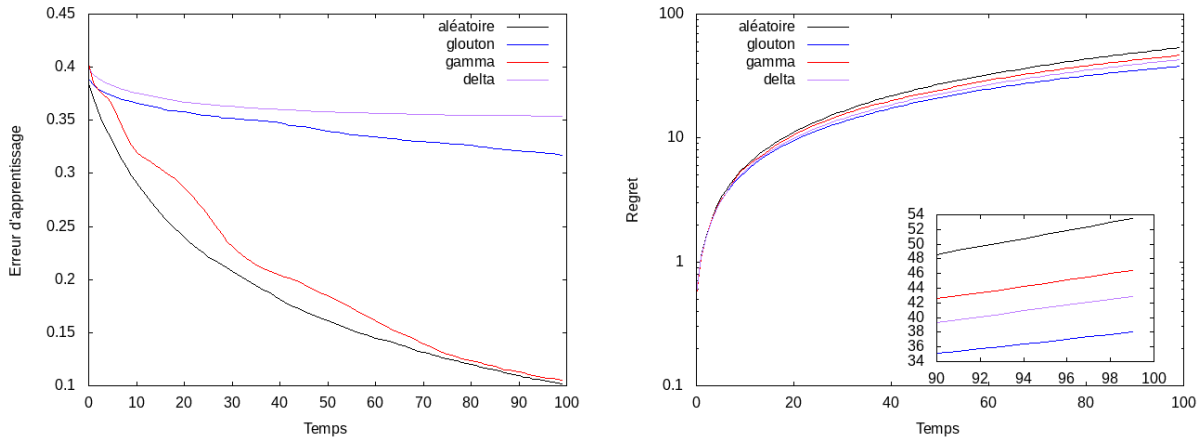
4.2.2.2 Analyse des résultats

Commençons par l'erreur d'apprentissage. Pour 5 agents, la stratégie qui apprend le moins est δ-cœur, suivie d'ϵ-glouton. Les stratégies aléatoire et γ-cœur quant à elles



(a) Erreur moyenne absolue (b) Regret cumulé

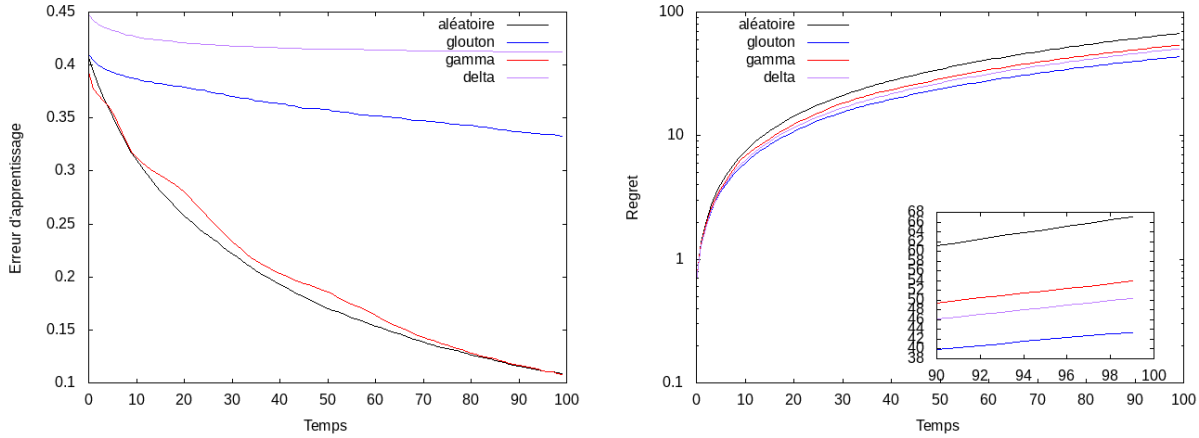
FIGURE 4.6 – Résultats pour 5 agents pour chaque stratégie avec la classe Normal



(a) Erreur moyenne absolue (b) Regret cumulé

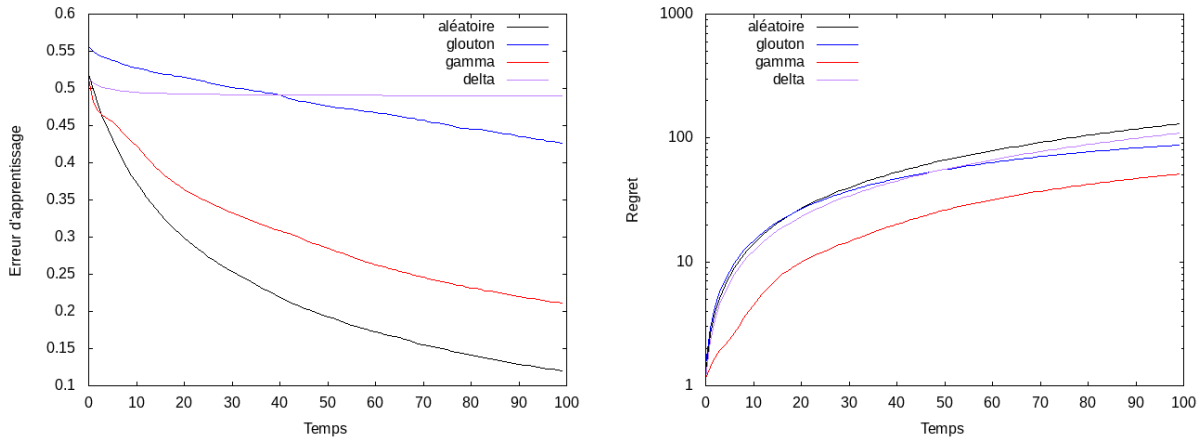
FIGURE 4.7 – Résultats pour 5 agents pour chaque stratégie avec la classe Uniform

apprennent beaucoup plus la fonction caractéristique réelle, et approximativement au même taux, sauf pour la classe *Random* où la stratégie γ-cœur apprend un peu moins bien. En revanche, dans ce dernier cas, la stratégie γ-cœur est celle qui possède le regret cumulé le plus bas (et donc le meilleur), où elle est suivie par ϵ-glouton, δ-cœur et aléatoire qui restent assez proches. Pour la classe *Normal*, ϵ-glouton rejoint γ-cœur en termes de regret minimal, bien que les quatre stratégies possèdent des regrets très proches. Pour les deux autres classes, à savoir *Uniform* et *NDCS*, les résultats sont les mêmes : ϵ-glouton est la meilleure stratégie, suivie de δ-cœur puis γ-cœur, et enfin la stratégie aléatoire. Pour 6 agents, l'erreur d'apprentissage est maximale pour δ-cœur avec toutes les classes,



(a) Erreur moyenne absolue (b) Regret cumulé

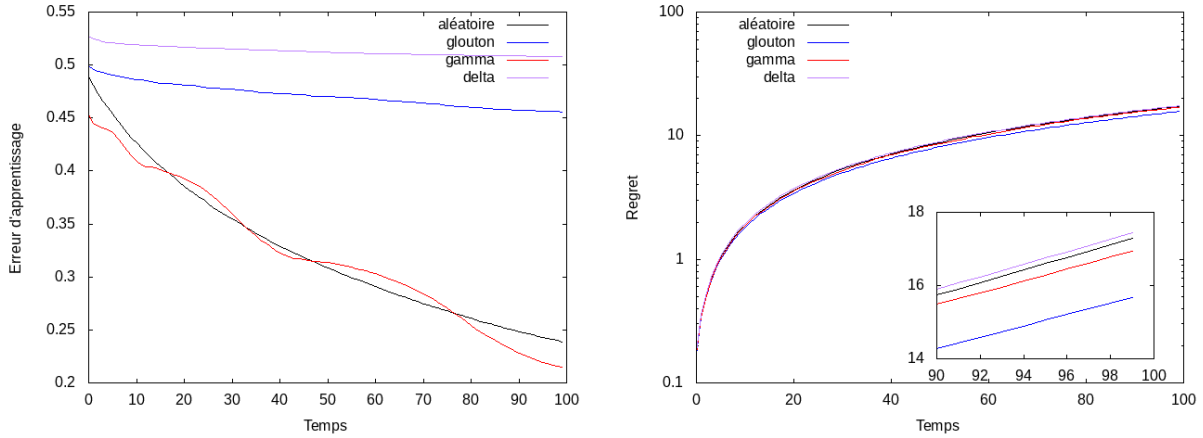
FIGURE 4.8 – Résultats pour 5 agents pour chaque stratégie avec la classe NDCS



(a) Erreur moyenne absolue (b) Regret cumulé

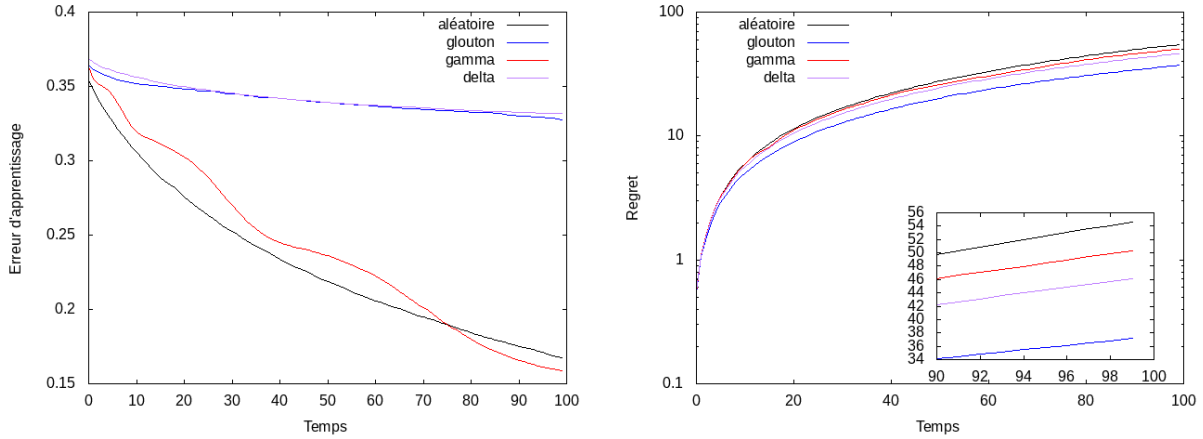
FIGURE 4.9 – Résultats pour 5 agents pour chaque stratégie avec la classe Random

mais est rejoint par ϵ-glouton pour les classes *Uniform* et *NDCS*. Pour les deux autres stratégies, nous sommes dans le même cas que pour 5 agents, à savoir qu'elles sont proches, sauf pour la classe *Random* où γ-cœur apprend moins bien. Cette dernière stratégie est également la meilleure pour cette classe en termes de regret cumulé, ce dernier suivant les mêmes schémas que pour 5 agents dans les différentes classes, c'est-à-dire que la stratégie ϵ-glouton est la meilleure, et δ-cœur la deuxième sauf pour la classe *Normal*. Cependant, cette dernière stratégie devient la meilleure en termes de regret dans la classe *Normal* pour 7 agents, bien que le regret de toutes les stratégies soit proche. Hormis pour la classe *Random*, c'est ϵ-glouton qui apprend le moins bien, suivie de δ-cœur, cette dernière



(a) Erreur moyenne absolue (b) Regret cumulé

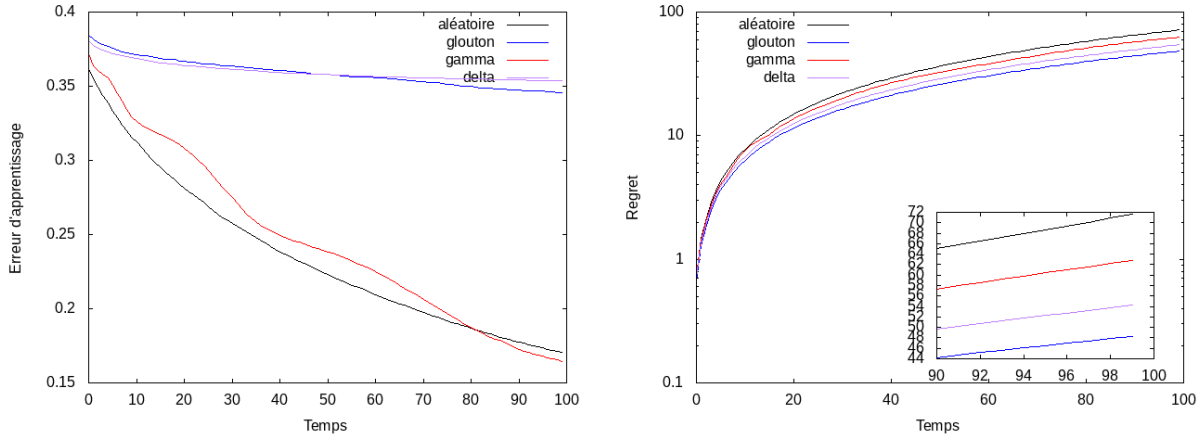
FIGURE 4.10 – Résultats pour 6 agents pour chaque stratégie avec la classe Normal



(a) Erreur moyenne absolue (b) Regret cumulé

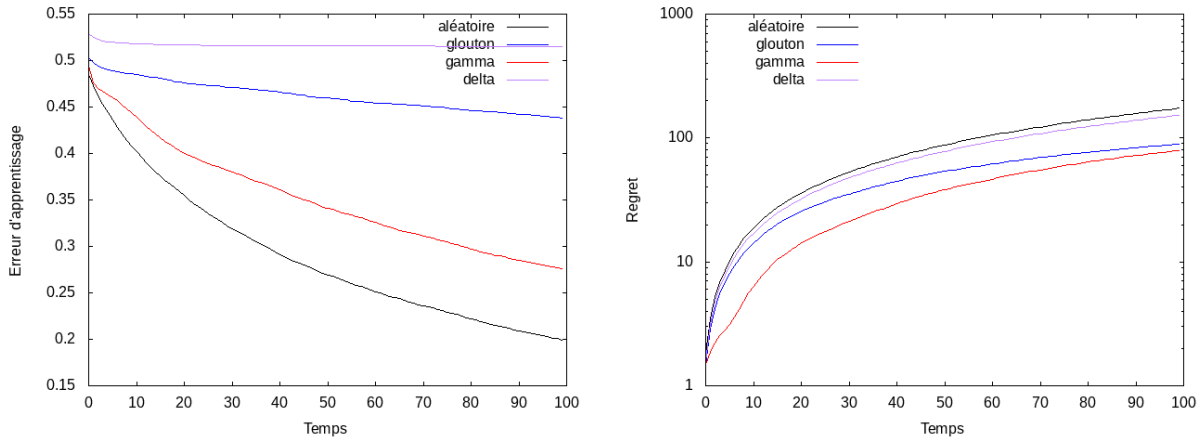
FIGURE 4.11 – Résultats pour 6 agents pour chaque stratégie avec la classe Uniform

dépassant légèrement la première dans le cas de la classe *Random*. Tout comme pour 5 et 6 agents, la stratégie γ-cœur apprend moins bien que la stratégie aléatoire sur la classe *Random*, mais c'est également le cas pour la classe *NDCS*. En revanche, elle apprend un peu plus la classe *Uniform*, tandis que pour la classe *Normal* les deux stratégies se valent. Les tendances du regret sont les mêmes que pour 6 agents en dehors du cas de la classe *Normal* évoqué précédemment, où δ-cœur devient la meilleure.



(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.12 – Résultats pour 6 agents pour chaque stratégie avec la classe NDCS

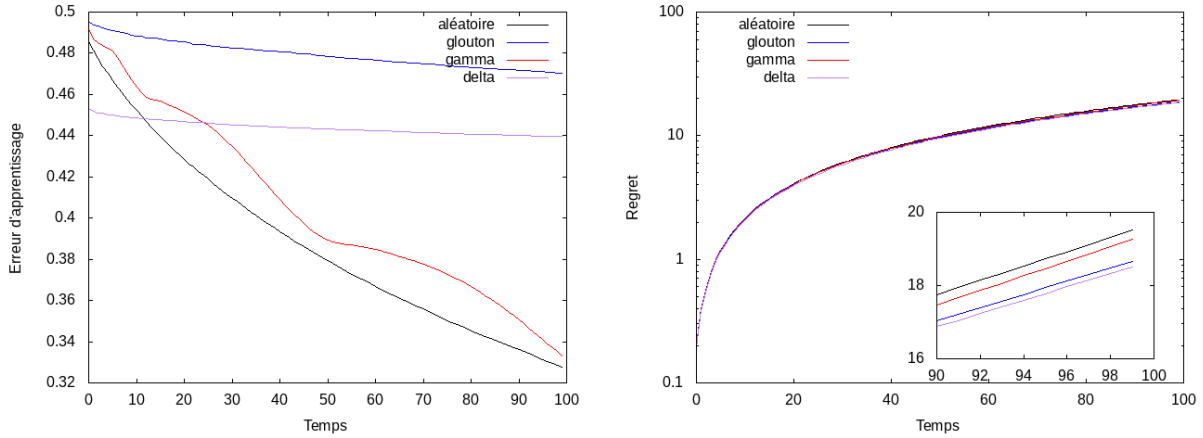


(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.13 – Résultats pour 6 agents pour chaque stratégie avec la classe Random

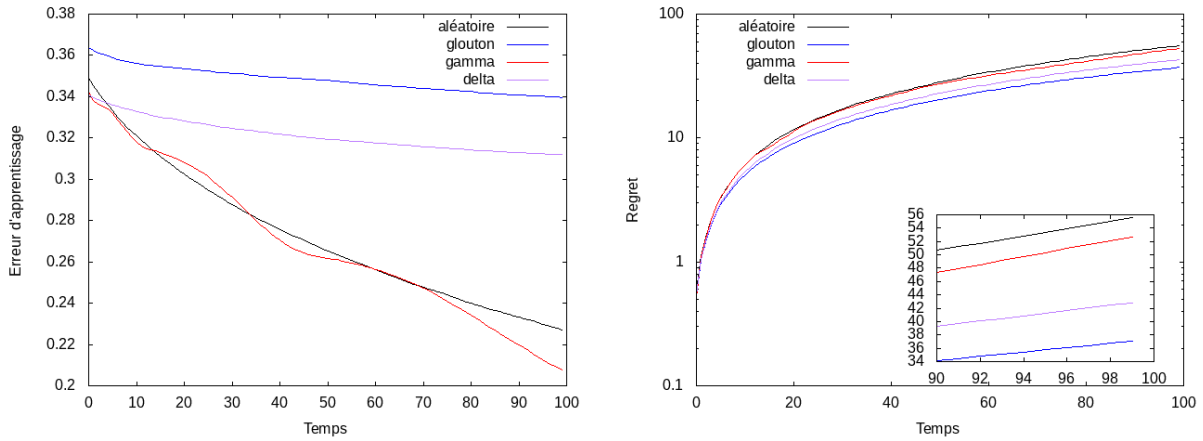
4.2.2.3 Conclusion intermédiaire

Nous pouvons commencer par conclure qu'un meilleur apprentissage de la fonction caractéristique n'entraîne pas un meilleur regret. En effet, si nous prenons en exemple les stratégies δ-cœur et ϵ-glouton, celles-ci sont celles qui ont la plus grande erreur d'apprentissage pour tous les paramétrages, et sont souvent celles possédant le regret minimal, en dehors du cas spécifique de la classe *Random*, où c'est γ-cœur qui s'illustre. D'un point de vue plus global, c'est la stratégie ϵ-glouton qui est donc la meilleure en termes de regret, ce qui est cohérent avec les résultats obtenus précédemment quant à l'effet de l'exploration,



(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.14 – Résultats pour 7 agents pour chaque stratégie avec la classe Normal



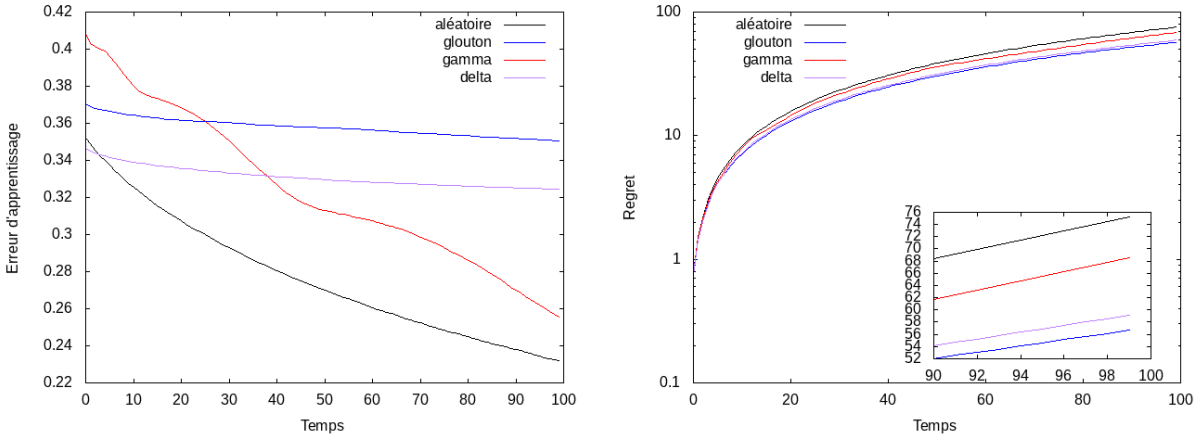
(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.15 – Résultats pour 7 agents pour chaque stratégie avec la classe Uniform

qui montraient que plus il y avait d'exploration, plus le regret était fort.

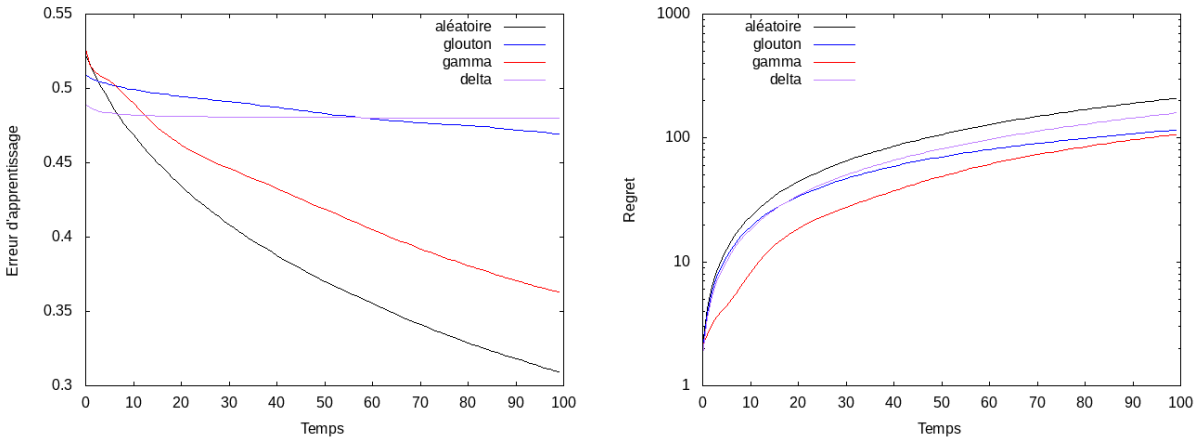
4.3 Expérimentations : estimation par inférence

Afin de déterminer l'impact de l'inférence sur les concepts fondés sur l'exploration, nous reproduisons exactement les mêmes expériences, mais avec un modèle d'estimation différent. Tout comme pour les expérimentations avec la méthode d'estimation fondée sur les connaissances *a priori*, nous commencerons par l'analyse des effets de l'exploration avec la stratégie ϵ-glouton, puis nous mettrons en concurrence à cette dernière nos concepts de



(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.16 – Résultats pour 7 agents pour chaque stratégie avec la classe NDCS



(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.17 – Résultats pour 7 agents pour chaque stratégie avec la classe Random

solutions fondés sur l'exploration, mais également avec une stratégie purement aléatoire. L'objectif ici est donc d'analyser si le fait d'inférer les connaissances permet une meilleure décision pour les agents.

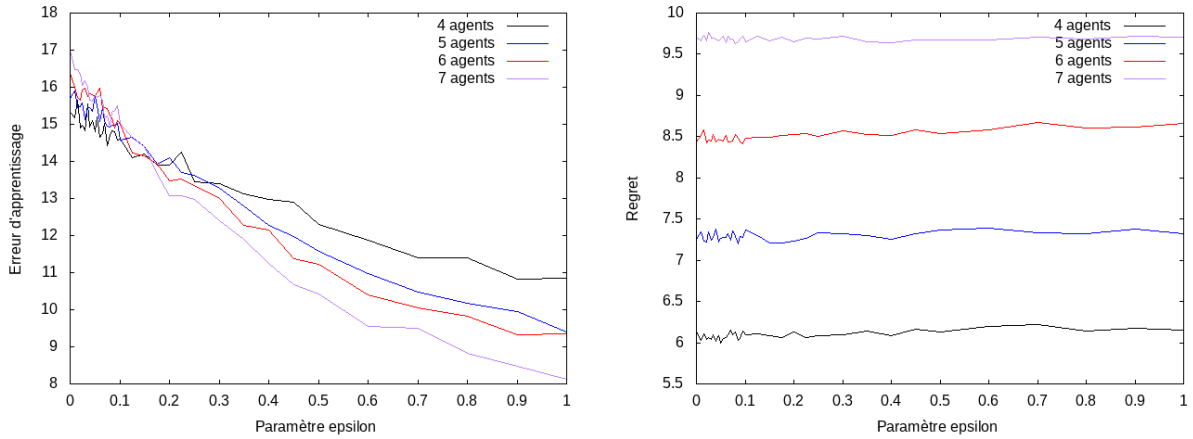
4.3.1 Effets de l'exploration avec ϵ-glouton

Tout comme précédemment, ces expérimentations ont pour objectif d'évaluer les performances de la stratégie ϵ-glouton dont le terme d'exploitation est remplacé par le concept de solutions ϵ-cœur, lorsque nous faisons varier la valeur ϵ de la stratégie, en allant de 0

à 1. De la même manière que pour cette expérience dans un cadre sans inférence, les pas de la valeur ϵ sont de plus en plus espacés à mesure que celle-ci grandit et nous générons également des jeux de 4 agents, paramétrés de la même manière que ceux de 5 à 7 agents.

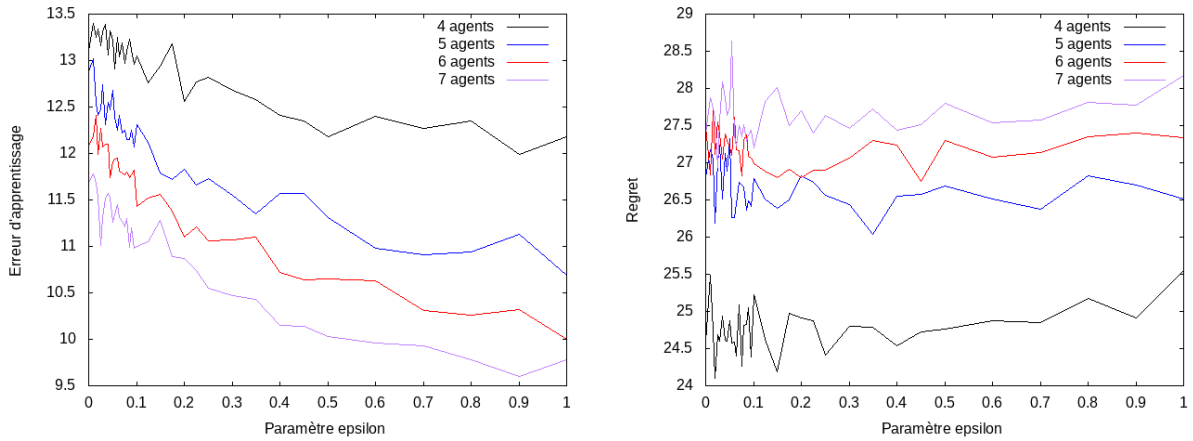
4.3.1.1 Graphiques

Les figures 4.18, 4.19, 4.20 et 4.21 présentent les résultats sur l'erreur d'apprentissage et le regret cumulé respectivement pour les classes *Normal*, *Uniform*, *NDCS*, et *Random*, et ce en fonction du nombre d'agents (allant de 4 à 7), avec en abscisse les valeurs d'ϵ pour la stratégie ϵ-glouton qui ont été testées.



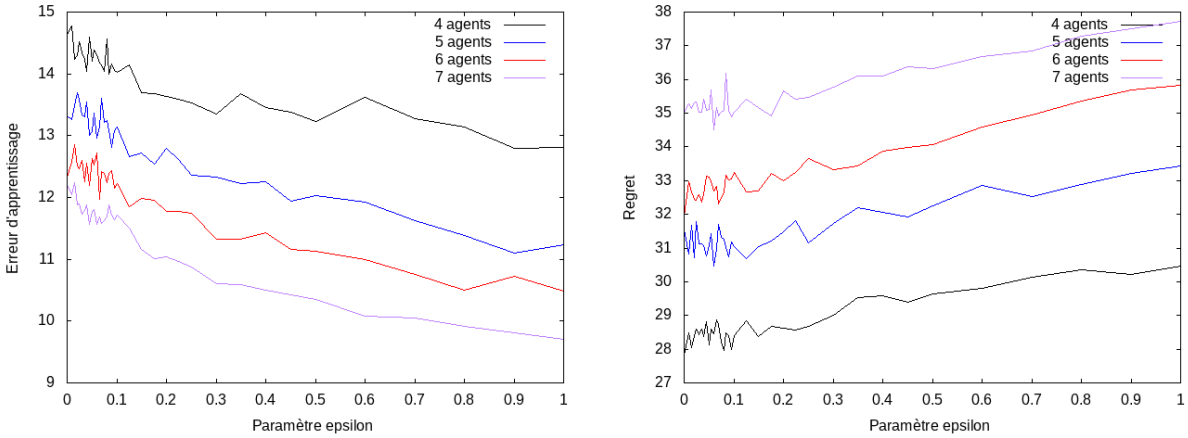
(a) Erreur moyenne absolue finale (b) Regret cumulé

FIGURE 4.18 – Résultats pour la stratégie ϵ-gloutonne avec la classe Normal



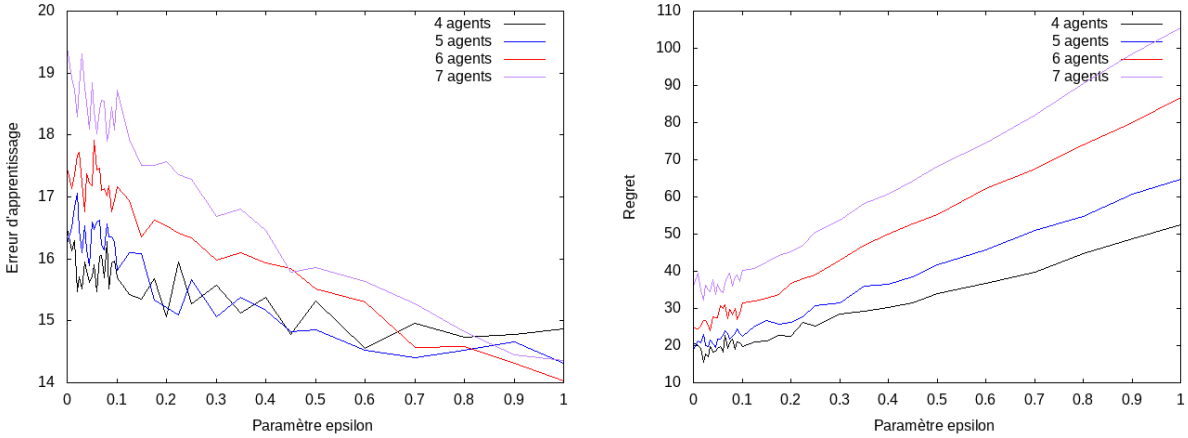
(a) Erreur moyenne absolue finale (b) Regret cumulé

FIGURE 4.19 – Résultats pour la stratégie ϵ-gloutonne avec la classe Uniform



(a) Erreur moyenne absolue finale (b) Regret cumulé

FIGURE 4.20 – Résultats pour la stratégie ϵ-gloutonne avec la classe NDCS



(a) Erreur moyenne absolue finale (b) Regret cumulé

FIGURE 4.21 – Résultats pour la stratégie ϵ-gloutonne avec la classe Random

4.3.1.2 Analyse des résultats

Pour analyser les résultats, il convient de savoir que les nombreuses oscillations, lorsque ϵ est petit, sont dues au plus grand nombre de données sur l'intervalle $[0, 0.1]$. Indépendemment du nombre d'agents, l'erreur d'apprentissage décroît lorsque ϵ croît, c'est-à-dire lorsque la stratégie dérive de très gloutonne à une exploration aléatoire, cela s'explique simplement par le fait que plus le terme d'exploration est fort, plus les coalitions sur lesquels il y a peu d'informations seront formées, ce qui entraîne une meilleure estimation de la part des agents. Toutefois, le regret reste le même ou augmente très légèrement

lorsque ϵ augmente également, en raison de l'exploitation plus faible lorsque ϵ augmente. Cependant, nous pouvons souligner le fait que même avec une exploration forte, le regret n'augmente pas drastiquement (sauf pour la classe *Random*), ce qui signifie que les solutions trouvées lors de l'exploration ne sont pas très éloignées de la solution optimale au sens de l'exploitation. Pour la classe *Random*, l'explication se trouve dans l'absence totale de structure, les structures de coalitions peuvent donc être bien plus hétérogènes en termes d'utilité. Ainsi, bien que l'exploration est intéressante afin de mieux estimer les fonctions caractéristiques, son effet est beaucoup moins intéressant concernant la minimisation du regret dans un contexte de formation de coalitions stochastique répétée, avec un gain minime lorsqu'il y en a, à moins que la fonction caractéristique ne soit pas structurée.

4.3.1.3 Conclusion intermédiaire

Comme attendu, lorsque le terme d'exploration de la stratégie ϵ-glouton prend du poids (c'est-à-dire quand ϵ augmente), la stratégie permet une meilleure estimation de la fonction caractéristique réelle. Cependant, cet apprentissage se fait au détriment du regret qui augmente également, ou dans le meilleur des cas reste le même. L'impact du nombre d'agents est clairement négligeable si nous regardons les tendances des courbes. Concernant la valeur ϵ de la stratégie ϵ-glouton dans les expérimentations suivantes, nous avons choisi de prendre une valeur arbitraire dans l'intervalle le plus intéressant à la vue des résultats, à savoir l'intervalle $[0, 0.1]$. Le choix s'est donc porté sur la valeur médiane de cet intervalle, tel que $\epsilon = 0.05$.

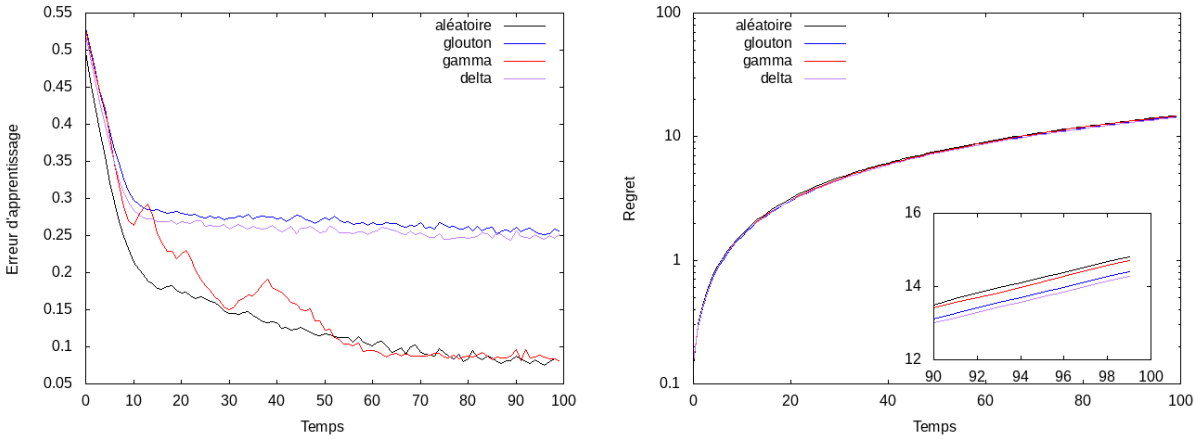
4.3.2 δ-cœur contre γ-cœur, ϵ-glouton et une décision aléatoire

Toujours avec la méthode d'estimation par inférence, nous reproduisons les expériences mettant en concurrence les deux concepts de solutions que nous proposons, avec également une stratégie aléatoire ainsi que la stratégie fondée sur l'ϵ-cœur présentée en section 4.1.2.2.

4.3.2.1 Graphiques

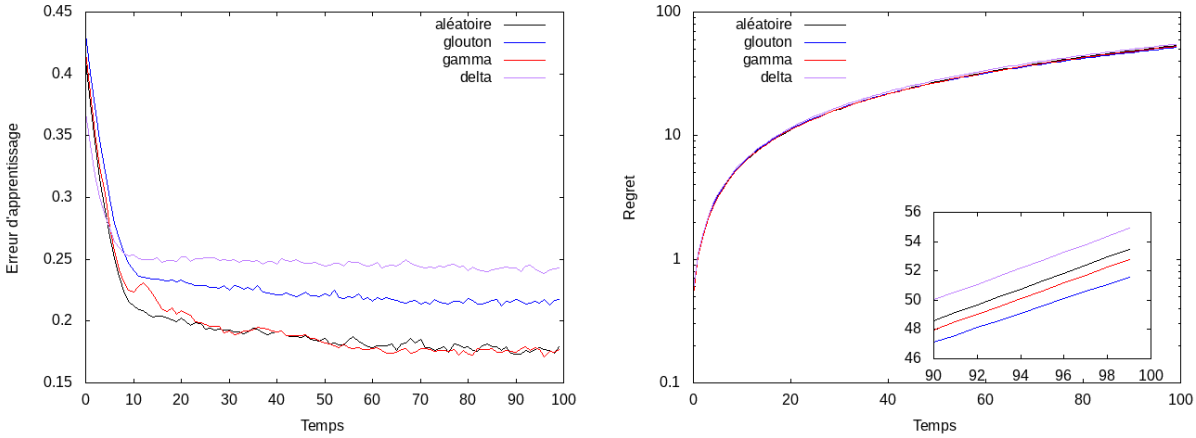
Les figures 4.22, 4.23, 4.24 et 4.25 représentent les expérimentations sur l'effet de l'exploration respectivement pour les classes *Normal*, *Uniform*, *NDCS* et *Random* pour 5 agents, les figures 4.26, 4.27, 4.28 et 4.29 celles pour les classes *Normal*, *Uniform*, *NDCS* et *Random* pour 6 agents, et enfin les figures 4.30, 4.31, 4.32 et 4.33 respectivement celles

pour les classes *Normal*, *Uniform*, *NDCS* et *Random* pour 7 agents. Afin de permettre une meilleur lisibilité des courbes de regret lorsque celles-ci sont très proches, un agrandissement sur les 10 derniers pas de temps est fourni en supplément pour les classes *Normal*, *Uniform* et *NDCS*. Pour la classe *Random*, les courbes sont distinguables les unes des autres.



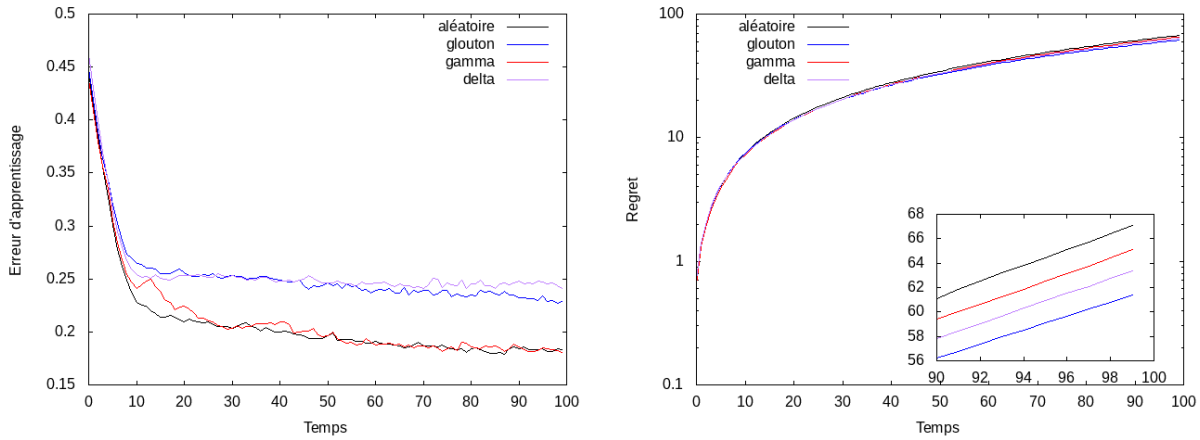
(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.22 – Résultats pour 5 agents pour chaque stratégie avec la classe Normal



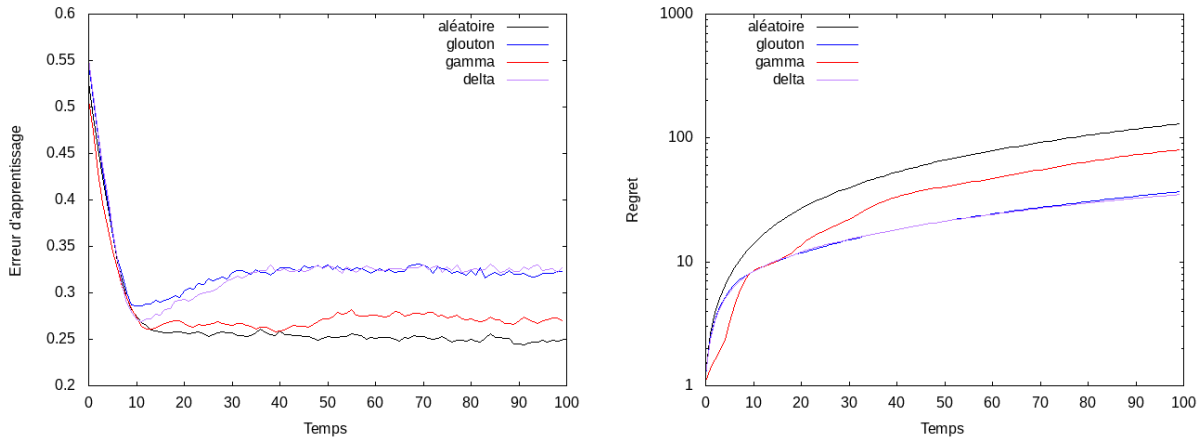
(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.23 – Résultats pour 5 agents pour chaque stratégie avec la classe Uniform



(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.24 – Résultats pour 5 agents pour chaque stratégie avec la classe NDCS

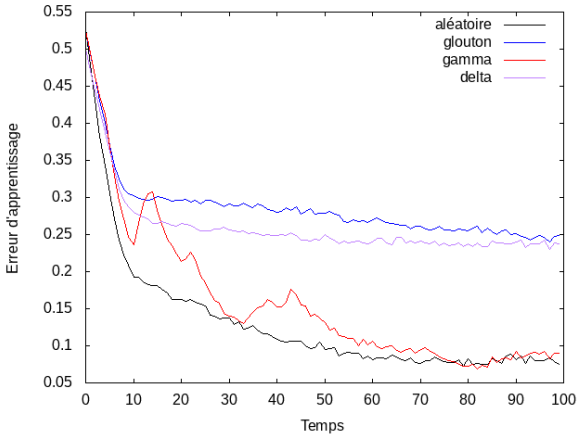


(a) Erreur moyenne absolue (b) Regret cumulé

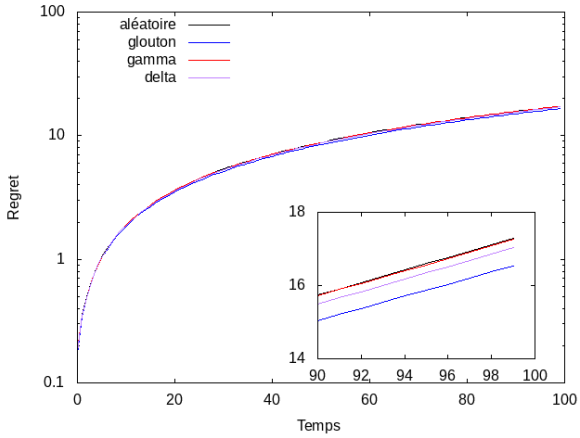
FIGURE 4.25 – Résultats pour 5 agents pour chaque stratégie avec la classe Random

4.3.2.2 Analyse des résultats

Un premier point pouvant être souligné concerne l'erreur d'apprentissage. Quelques soient le nombre d'agents et la classe de fonctions caractéristique utilisés, le même motif peut être observé à chaque fois : les stratégies γ-cœur et aléatoire ont un taux d'apprentissage extrêmement proche à la fin des 100 pas de temps, hormis pour la classe *Random* où la différence est un peu plus importante. Il peut arriver que dans les premiers pas de temps (entre 10 et 30 environ) les deux stratégies soient éloignées, notamment dû à un regain d'erreur de la part de la stratégie γ-cœur, mais cette dernière parvient à abaisser

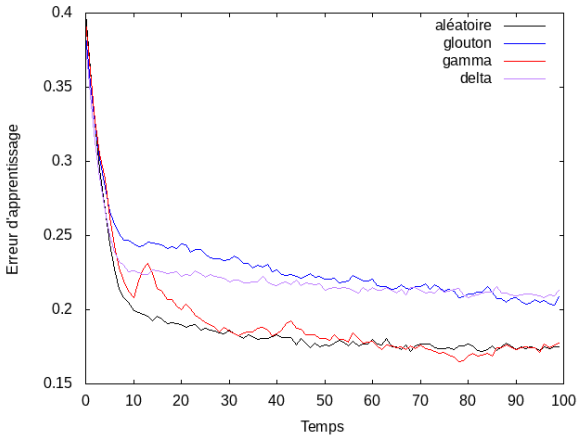


(a) Erreur moyenne absolue

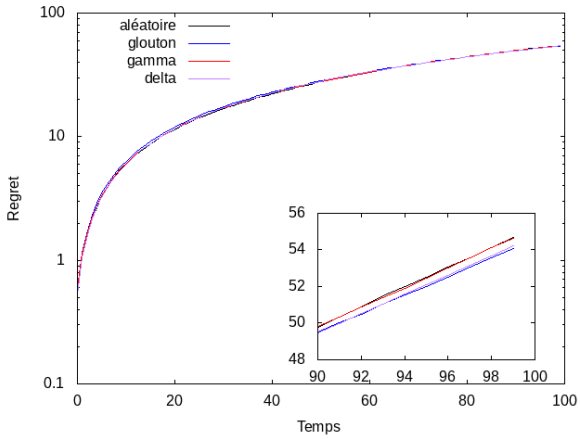


(b) Regret cumulé

FIGURE 4.26 – Résultats pour 6 agents pour chaque stratégie avec la classe Normal



(a) Erreur moyenne absolue

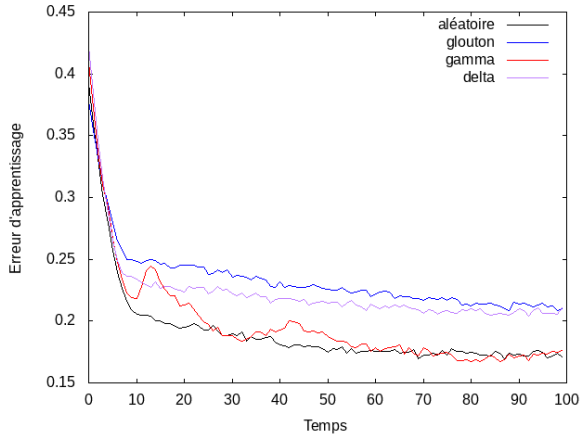


(b) Regret cumulé

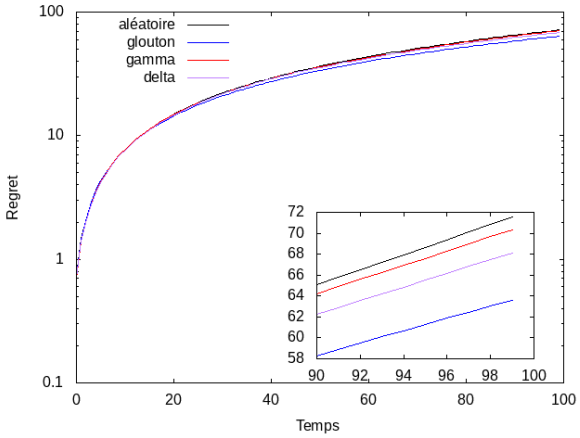
FIGURE 4.27 – Résultats pour 6 agents pour chaque stratégie avec la classe Uniform

cette erreur au niveau de la stratégie aléatoire à chaque fois, en témoignent notamment les expérimentations sur les fonctions caractéristiques structurées par la classe *Normal*. Toujours à propos de l'erreur d'apprentissage, la 3ème meilleure stratégie permettant le meilleur apprentissage après γ-cœur et aléatoire est δ-cœur. En effet, δ-cœur est majoritairement meilleure que ϵ-glouton sur cette mesure, aux trois exceptions près des classes *Uniform*, *NDCS* pour 5 agents, et *Random* pour 6 agents, où les deux stratégies ont une erreur d'apprentissage similaire.

Concernant le regret cumulé, nous pouvons observer les tendances exactement inverses : la stratégie majoritairement meilleure que les autres est ϵ-glouton. Hors exception,

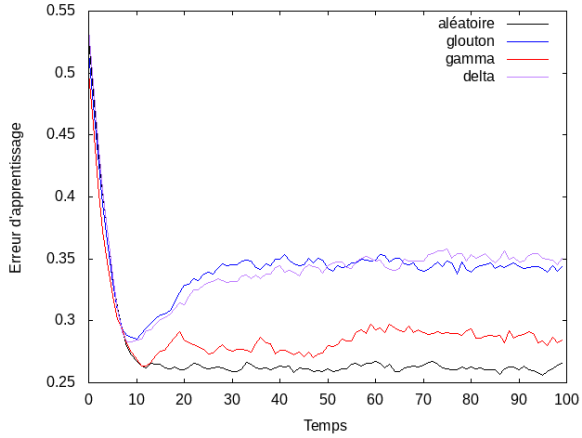


(a) Erreur moyenne absolue

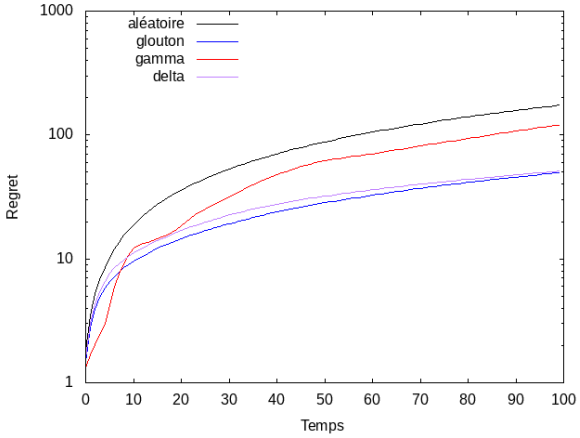


(b) Regret cumulé

FIGURE 4.28 – Résultats pour 6 agents pour chaque stratégie avec la classe NDCS



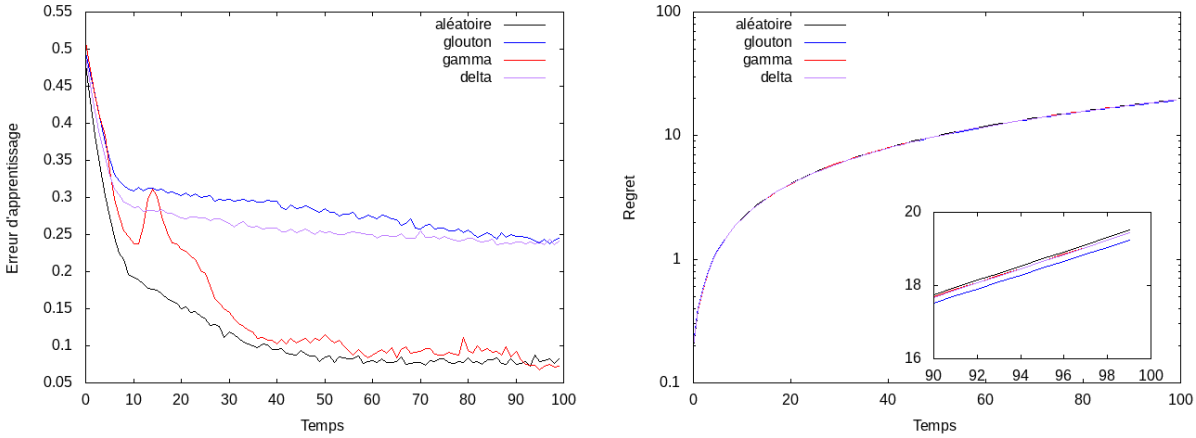
(a) Erreur moyenne absolue



(b) Regret cumulé

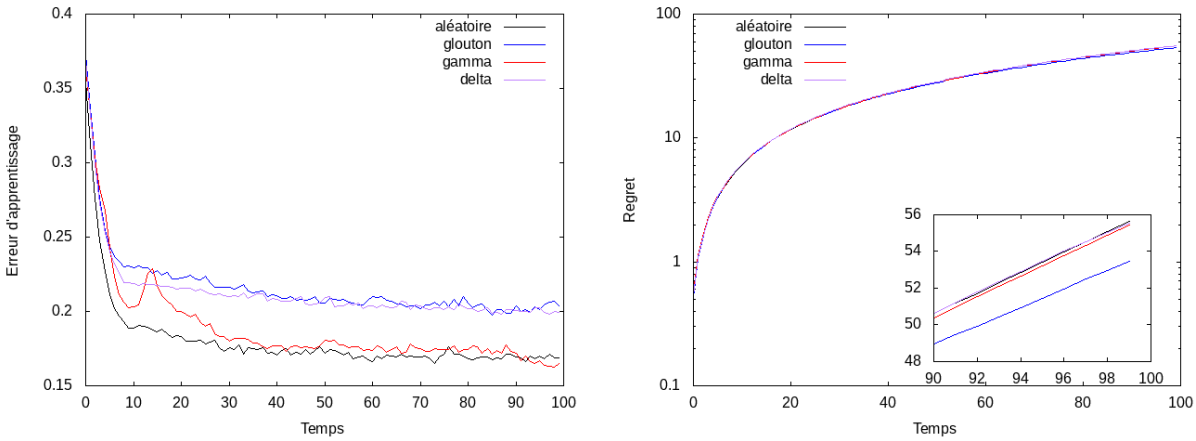
FIGURE 4.29 – Résultats pour 6 agents pour chaque stratégie avec la classe Random

c'est la stratégie qui atteint en général le regret cumulé le plus bas parmi les stratégies, suivie de δ-cœur (hormis pour la classe *Uniform* pour 5 agents où elle arrive dernière). Dans plusieurs cas de figures (*Normal* pour 5, 6 et 7 agents, *Random* pour 5 et 7 agents), δ-cœur est très proche ou meilleur que ϵ-glouton. Les stratégies γ-cœur et aléatoire sont souvent confondues dans ces résultats, avec un léger avantage à la première lorsque ce n'est pas le cas. La classe de fonctions caractéristiques donnant les résultats les moins clairs est la classe *Normal*, quelque soit le nombre d'agents, tandis que la classe où cela est le plus visible est *Random*, également pour tout nombre d'agents.



(a) Erreur moyenne absolue (b) Regret cumulé

FIGURE 4.30 – Résultats pour 7 agents pour chaque stratégie avec la classe Normal

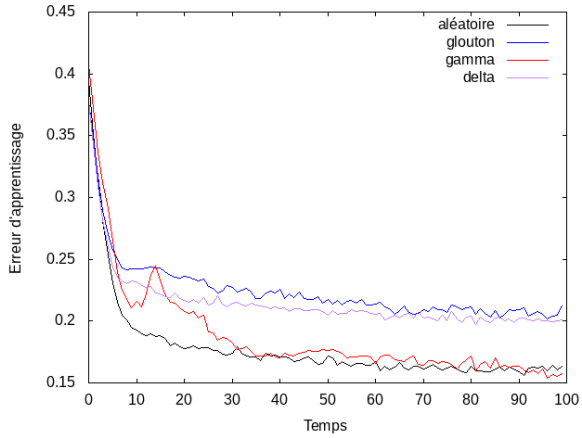


(a) Erreur moyenne absolue (b) Regret cumulé

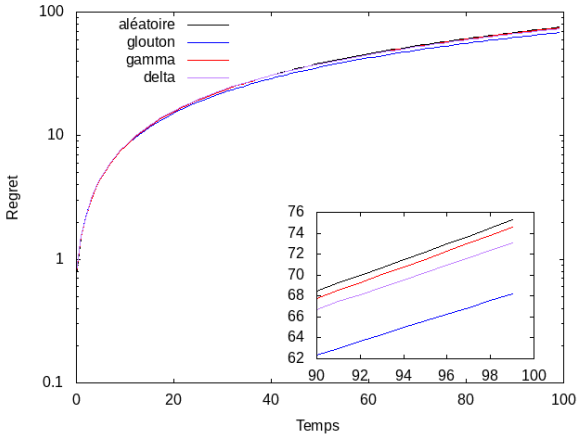
FIGURE 4.31 – Résultats pour 7 agents pour chaque stratégie avec la classe Uniform

4.3.2.3 Conclusion intermédiaire

Au vu des résultats, nous pouvons conclure que dans un cadre où l'inférence est permise, une stratégie gloutonne intégrant une notion d'exploration simple telle que la stratégie ϵ-glouton suffit amplement. En effet, malgré une erreur d'apprentissage manisfestement plus grande, cette stratégie obtient un regret cumulé moindre que les autres stratégies. De tels résultats peuvent être expliqués par le fait que l'apprentissage précis de la fonction caractéristique n'est pas nécessaire, tant que la stratégie parvient à obtenir juste assez d'informations pour classer les structures de coalitions entre elles, sans avoir

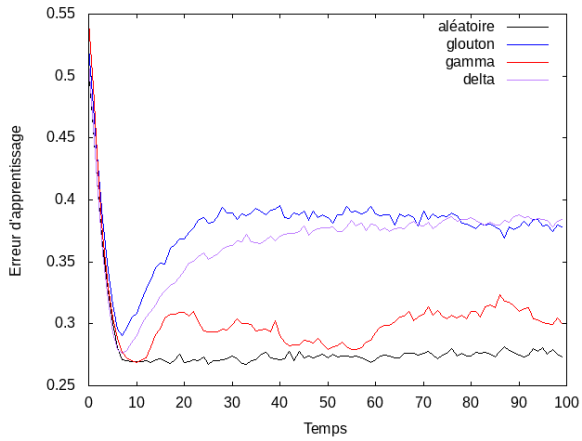


(a) Erreur moyenne absolue

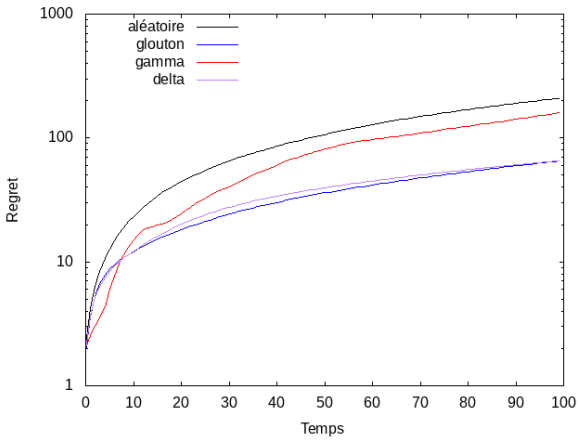


(b) Regret cumulé

FIGURE 4.32 – Résultats pour 7 agents pour chaque stratégie avec la classe NDCS



(a) Erreur moyenne absolue



(b) Regret cumulé

FIGURE 4.33 – Résultats pour 7 agents pour chaque stratégie avec la classe Random

une estimation pointilleuse de l'espérance de chaque coalition. De plus, si nous pouvons conclure que δ-cœur n'est pas aussi efficace que ϵ-glouton, nous pouvons également dire qu'elle est néanmoins plus efficace que γ-cœur, qui semble accorder un poids largement trop important à l'exploration. Les principaux points à retenir sont donc les suivants.

— Mieux apprendre la fonction caractéristique n'entraîne pas un meilleur regret,
— Les stratégies ont principalement besoin de classer les coalitions entre elles,
— δ-cœur est plus efficace que γ-cœur,
— δ-cœur est un peu moins efficace que ϵ-glouton, mais rivalise parfois.

4.4 Conclusion

Un premier point important à souligner est que l'apprentissage de la fonction caractéristique n'influence en rien les performances en termes de regret. En effet, bien qu'avec un modèle d'estimation fondé sur une connaissance *a priori*, le regret est maximal lorsque l'erreur d'apprentissage est maximale, cela n'est pas le cas avec un modèle d'estimation fondé sur l'inférence, où lorsque nous avons une stratégie purement exploratoire, l'erreur d'apprentissage est plus faible, mais le regret reste toujours plus grand que par rapport à une stratégie purement gloutonne. Cette analyse est renforcée par le fait que les deux stratégies les plus performantes en termes de regret, que ce soit avec ou sans inférence, sont δ-cœur et ϵ-glouton, qui sont également celles qui apprennent le moins bien. Le comportement de la stratégie γ-cœur va également dans ce sens, car devient la meilleur en termes de regret avec les fonctions caractéristiques de la classe *Random* lorsque nous utilisons une méthode d'estimation par inférence, tandis qu'elle apprend moins bien dans cette configuration. Une différence notable sur l'effet de l'exploration pour les différentes méthodes d'estimation est à souligner. En effet, la méthode par inférence profite d'une plus grande part d'exploration afin de réduire son erreur d'apprentissage, tandis que l'inverse se produit pour la méthode fondée sur une connaissance *a priori*. Cet effet provient principalement de l'initialisation de la connaissance dans ces méthodes, qui est déterminée de façon aléatoire selon les poids du réseau de neurones pour la méthode par inférence, et un tirage aléatoire suivant la même classe de structure que la fonction caractéristique réelle pour l'autre méthode. Cela a pour conséquence que les méthodes ne partent pas du tout du même *a priori*, qui est très mauvais pour la méthode par inférence, et correct pour l'autre. Enfin, pour les deux méthodes d'estimations, nous pouvons conclure que la stratégie ϵ-glouton est la plus perfomante au global, bien que δ-cœur ne soit pas très loin derrière, et la dépasse parfois dans le cadre de l'estimation par inférence. En revanche, la stratégie γ-cœur est quant à elle souvent au niveau d'une stratégie aléatoire.

Chapitre 5

Protocole déterministe de formation de coalitions basé sur des concessions

Sommaire

Comme présenté au début de ce mémoire, nos problématiques portent sur trois hypothèses communément faites dans la formation de coalitions. Nous avons abordé deux d'entre elles dans les deux chapitres précédents, à savoir le déterminisme et la connaissance *a priori* des utilités des coalitions. La troisième et dernière problématique est donc

la décentralisation du problème de formation de coalitions. Bien que ce problème ait déjà été abordé dans la littérature, les algorithmes proposés sont bien souvent conçus pour correspondre à un contexte spécifique, en faisant par exemple des hypothèses fortes sur le modèle sous-jacent. Notre objectif est donc de proposer un processus de formation de coalitions décentralisé en nous abstrayant de tout contexte, ainsi qu'en évitant les éléments simulant une centralisation, comme c'est par exemple le cas avec les commissaires-priseurs dans les approches par négociation. Nous proposons ainsi d'adapter au problème de la formation de coalitions un protocole de concessions monotones pour la négociation multilatérale entre agents utilitaristes proposé originellement par Ulle Endriss [Endriss, 2006]. Ce protocole est exempt d'hypothèse sur la structure du système et sur la présence d'entité centrale. Il est donc un bon candidat pour répondre à notre problématique.

Dans ce chapitre, nous redéfinissons certains éléments de ce protocole, et proposons de nouvelles stratégies afin de prendre en compte la notion de coalitions. Dans un premier temps, nous proposons une version distribuée du protocole, avant d'y apporter des modifications pour le décentraliser complètement dans un second temps. Enfin, une étude empirique est menée afin de mesurer l'intérêt de ces deux protocoles l'un par rapport à l'autre, et également par rapport à une solution centralisée.

5.1 Un protocole de concessions distribué

Dans le cadre de la négociation multilatérale entre agents utilitaristes, Ulle Endriss a proposé en 2006 un protocole intéressant [Endriss, 2006]. Son intérêt réside dans le fait qu'il ne possède pas d'entité centrale, qu'il est distribué, et qu'il ne fait pas d'hypothèse sur la structure du système. Dans un premier temps, nous présentons ce protocole, puis nous établissons les liens que nous pouvons faire entre ce dernier et la formation de coalitions. Enfin, grâce à ces liens, nous proposons un nouveau protocole de concessions monotones distribué adapté à la formation de coalitions.

5.1.1 Un protocole de concessions monotones

L'objectif du protocole proposé par Ulle Endriss [Endriss, 2006] est de permettre à un ensemble d'agents égoïstes de négocier dans le but d'obtenir le meilleur gain possible. Pour cela, les agents formulent chacun des propositions et doivent se mettre d'accord sur une proposition commune. La négociation se déroule en plusieurs tours, en suivant un protocole

de *concessions monotones*. La notion de *concession* décrit le fait qu'un agent accepte d'abandonner sa proposition afin d'en formuler une nouvelle dans l'objectif que cela mène à un accord. En faisant cela, nous disons que l'agent *concède*. Ces concessions sont dites *monotones* car les agents font en premier lieu des propositions qui sont particulièrement bénéfiques pour eux-mêmes, puis ils révisent itérativement leur proposition précédente (lors des concessions) afin d'éventuellement arriver à un accord, en réduisant à chaque étape le gain qu'ils s'accordent ou en proposant davantage aux autres agents.

5.1.1.1 Propositions

Au premier tour, chaque agent fait une proposition initiale, en proposant celle qui lui rapporte le gain le plus élevé. Une fois les propositions initiales choisies, les agents peuvent maintenir leur proposition ou concéder. Ainsi, à chaque tour suivant, les agents qui concèdent font de nouvelles propositions, à nouveau proposées simultanément avec celles gardées par les autres agents, puis les agents décident de maintenir leur proposition ou concéder, et ainsi de suite jusqu'à ce qu'un accord commun ou un conflit (aucun agent ne peut concéder) émerge. Les propositions sont choisies parmi un ensemble fini de propositions possibles.

Définition 5.1 (Ensemble des propositions)**.** *Soit N un ensemble fini d'agents. L'ensemble $\mathcal{X}$ est un ensemble de propositions possibles, tel que pour chaque agent $a_i \in N$, il existe une fonction d'utilité $x_i : \mathcal{X} \mapsto \mathbb{R}_0^+$ qui associe le gain (non-négatif) de a_i à chaque proposition. $\mathcal{X}$ contient toujours une proposition p^0 où tous les agents gagnent* 0.

Notons que la proposition p^0 représente l'échec des négociations, qui permettra par la suite de définir la notion de *conflit*. Les propositions appartenant à l'ensemble des propositions sont définies comme suit.

Définition 5.2 (Proposition)**.** *Une* proposition $p_i = \{x_1, \ldots, x_n\}$ *de l'agent a_i est un vecteur de gain où $x_j(p_i) \geq 0$ est le gain de l'agent a_j. Un agent a_i* préfère *une proposition p à une proposition p' – noté $p \succeq_i p'$ – si, et seulement si, $x_i(p) \geq x_i(p')$.*

Il est donc fait l'hypothèse que les agents préfèrent les propositions où leur gain est le plus élevé. Un exemple de propositions dans un cadre de négociation multilatérale est donné dans le tableau 5.1 de l'exemple 28 sur la page 111.

5.1.1.2 Accord et conflit

Les accords communs et les conflits sont définis comme suit.

Définition 5.3 (Accord commun)**.** *Un* accord commun *est atteint si, et seulement si, un agent fait une proposition que tous les agents préfèrent à leur propre proposition. Formellement,*

$$\exists\ a_i \in N\ t.q.\ p_i \succeq_j p_j\ \forall\ a_j \in N$$

Nous pouvons remarquer que les propositions des agents n'ont pas besoin d'être identiques pour qu'il y ait un accord. En effet, si un agent fait une proposition dans laquelle le gain de chaque agent est supérieur à ce qu'ils se proposent à eux-mêmes, alors l'accord est atteint.

Définition 5.4 (Conflit)**.** *Un* conflit *émerge lorsqu'à un pas de temps, il n'y a pas d'accord commun et qu'aucun agent ne peut concéder, c'est-à-dire qu'aucune nouvelle proposition ne peut être formulée. La proposition p^0 constituée de gains nuls pour tous les agents est alors donnée comme résultat de la négociation.*

5.1.1.3 Stratégies et types de concession

Ensuite, la deuxième étape s'appuie sur deux notions importantes : la *stratégie de concession* et le *type de concession*. La première décrit comment les agents décident qui doit concéder, et la seconde décrit comment les concessions (i.e. les nouvelles propositions) doivent être faites. Dans son travail, Endriss présente deux stratégies. La première est *Willingness to Risk Conflict* (WRC) et la seconde est *Product Increasing* (PI). Ces stratégies s'appuient sur une valeur de Zeuthen, aussi appelée dans la suite valeur Z, qui quantifie l'intérêt d'un agent à concéder.

Pour la stratégie WRC, la valeur Z calculée décrit le risque que prend un agent s'il concède au regard de ce que les autres agents lui offrent. La stratégie stipule donc que l'agent devant concéder est celui qui prend le moins de risque. Informellement, l'agent qui concède est alors celui qui perdra le moins en retirant sa proposition.

Définition 5.5 (Willingness to Risk Conflict)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \begin{cases} 1 & si\ \ x_i(p_i) = 0, \\ \frac{x_i(p_i) - \min\limits_{\forall j \in N} x_i(p_j)}{x_i(p_i)} & sinon. \end{cases}$$

Quant à la stratégie PI, la valeur Z calculée décrit à quel point les gains des agents sont égaux dans une proposition. Plus ceux-ci sont différents, moins la valeur sera haute. Ainsi, cette stratégie favorise une distribution égalitaire, c'est-à-dire que les agents qui distribuent le moins également les gains seront les premiers à concéder.

Définition 5.6 (Product Increasing)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z'_{a_i} est la plus petite, où :*

$$Z'_{a_i} = \prod_{\forall j \in N} x_j(p_i)$$

Un exemple d'application des stratégies sur des propositions est donné dans le tableau 5.1 de l'exemple 28.

Exemple 28. *Soit l'ensemble d'agents $N =$* {  ,  ,  }*, un exemple de propositions et de calcul des valeurs de concession est le suivant. L'agent*  *concédera avec la stratégie WRC, tandis que pour la stratégie PI, ce sera l'agent*  *.*

	x 	x 	x 	WRC	PI
p 	2.60	2.12	1.81	<u>0.18</u>	9.98
p 	2.12	3.33	0.87	0.68	6.40
p 	2.38	1.06	2.47	0.65	<u>6.23</u>

TABLE 5.1 – Exemple de propositions et de valeurs de concessions

La seconde notion importante est celle de type de concession. Ces types décrivent les différentes façons que les agents concédants ont de construire une nouvelle proposition afin qu'elle soit acceptable pour les autres. Ces types représentent une contrainte sur la différence entre le vecteur de gain de l'ancienne proposition et celui de la nouvelle. Les différents types sont les suivants :

- **Fort** : le gain de tous les autres agents croît,
- **Faible** : le gain d'un autre agent croît,
- **Pareto** : le gain des agents reste au moins égal, et croît pour au moins un,
- **Égalitaire** : le gain minimal des autres agents croît,
- **Utilitaire** : la somme des gains des autres agents croît,
- **Nash** : le produit des gains des autres agents croît,
- **Égocentrique** : le gain de l'agent concédant décroît.

Exemple 29. *Reprenons les propositions de l'exemple 28. La notation $p_i \rightarrow p_j$ désigne le fait que l'agent a_i concède et fait la même proposition que a_j au pas de temps suivant. Les types de concession respectés par les propositions sont montrés dans le tableau 5.2. Remar-*

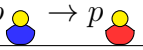
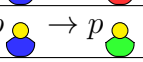
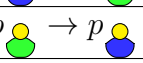
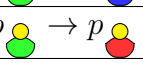
$p_i \rightarrow p_j$	Fort	Fai.	Par.	Éga.	Uti.	Nash	Égo.
	✗	✓	✗	✗	✓	✗	✓
	✗	✓	✗	✗	✗	✗	✓
	✓	✓	✓	✓	✓	✓	✓
	✗	✓	✗	✓	✓	✓	✓

TABLE 5.2 – Exemples de concessions typées

quons dans cet exemple que les agents formulent leurs nouvelles propositions comme étant celles d'autres agents. Ceci n'a que pour but d'illustrer les différents types de concession, les nouvelles propositions ne sont pas contraintes dans le protocole à se ramener à celles déjà en jeu.

5.1.2 Formation de coalitions et protocole de concessions

Un parallèle peut être fait entre l'*accord commun* dans le protocole de négociation et la stabilité dans la formation de coalitions. Chacun de ces termes caractérise le moment où les agents sont tous d'accord. D'un autre côté, les concessions représentent des agents qui acceptent un gain plus faible afin d'atteindre la stabilité, comme le concept de l'ϵ-cœur qui autorise d'abandonner une part de gains pour éviter les déviations. Malgré ce point commun, nous pouvons nous demander comment adapter ce protocole à notre problème de formation de coalitions.

5.1.2.1 Notions à adapter

Le protocole proposé par Endriss étant un protocole de négociations multilatérale pour des agents individuels. Or, cela ne correspond pas au problème de formation de coalitions où les agents doivent raisonner en termes de groupes. Une première étape pour adapter le protocole est donc de redéfinir les propositions afin qu'elles prennent en compte les notions de coalitions et de structures de coalitions. Cependant, en raison du fait que nos jeux de coalitions sont à utilité transférable, ceux-ci acceptent un nombre infini d'imputations possibles. Il nous est donc impossible de définir un ensemble de propositions finies. Par

conséquent, nous proposerons de fixer la manière que les agents ont de construire les propositions. Les propositions étant donc différentes, nous devons aussi redéfinir ce qu'est un *accord commun* dans le cadre de la formation de coalitions, c'est-à-dire que celui-ci ne doit pas seulement prendre en compte les gains, mais également les structures de coalitions.

De plus, le travail d'Ulle Endriss ne fait pas mention de la manière dont l'ensemble des propositions est construit. Une deuxième étape pour adapter le protocole est donc de définir comment les agents construisent les propositions, et ce en prenant également en compte les notions de coalitions et d'utilité transférable, qui fait que les agents doivent s'accorder sur une distribution de l'utilité au sein des coalitions. Pour ce dernier point, diverses répartitions peuvent être envisagées, comme une répartition égalitaire (part égale, valeur de solidarité) ou bien équitable (valeur de Shapley ou Banzhaf). Nous renvoyons le lecteur à la section 1.2.2.2 pour une description des approches évoquées ici.

Enfin, concernant les stratégies présentées par Endriss, celles-ci sont appliquées à partir du vecteur de gain. Il est donc possible de ne pas les adapter. Cependant, les stratégies WRC et PI décrivent respectivement un risque de perte individuel et un équilibre parmi les gains. Ces deux visions ne sont pas adaptées à la formation de coalitions, la première car le gain individuel des agents dépend de la coalition dans lesquelles ils se trouvent, et la deuxième car la stabilité d'une structure ne dépend pas d'un gain totalement équitable parmi les agents qui peuvent être dans des coalitions diverses. Elle doivent donc être adaptées afin de prendre en compte la notion de structure de coalitions.

5.1.2.2 Notions inchangées

Lorsqu'un agent concède, il doit construire une nouvelle proposition en accord avec un des types de concession proposés par Endriss. De manière intéressante, nous n'avons pas besoin de les adapter dans le contexte de la formation de coalitions. En effet, les concepts sous-jacents de ces types de concession s'appliquent très bien aux vecteurs de gains.

Exemple 30. *Considérons le type de concession* Pareto. *Un agent devant faire une nouvelle proposition avec ce type ne doit pas faire baisser les gains des autres agents et doit faire augmenter le gain d'au moins un autre agent par rapport à l'ancienne proposition. Cela fait sens dans le contexte de la formation de coalitions.*

Un autre élément du protocole d'Endriss que nous n'avons pas besoin de redéfinir est celui de *conflit*. En effet, même dans un contexte de formation de coalitions, un conflit

n'arrivera que lorsqu'aucun agent ne pourra concéder. Comme cela a été indiqué précédemment, les stratégies originelles peuvent être également être conservées telles quelles.

5.1.3 Protocole de concessions pour la formation de coalitions

Nous proposons désormais de donner de nouvelles définitions aux éléments identifiés précédemment comme étant à adapter.

5.1.3.1 Propositions

Le premier élément à adapter est donc ce qu'est une proposition dans le cadre de la formation de coalitions. Dans les travaux d'Ulle Endriss, les agents négocient des gains individuels, tandis que dans la formation de coalitions, les agents doivent pouvoir raisonner sur des coalitions, mais aussi sur leurs gains individuels dans les coalitions. Il est donc naturel de définir une proposition comme étant une solution à un jeu de coalitions : une structure de coalitions et un vecteur de gains.

Définition 5.7 (Proposition)**.** *Étant donné un jeu $\mathcal{G}$, une* proposition *de l'agent a_i, notée p_i, est une solution $S_\mathcal{G} = \langle \mathcal{CS}, \vec{x} \rangle$ où $\vec{x}$ est un vecteur de gains $\langle x_j^{C_j^i} \rangle$ où C_j^i est la coalition de l'agent a_j dans la proposition p_i.*

Étant donné que nous sommes dans un cadre de jeux de coalitions à utilité transférable, nous ne pouvons pas définir d'ensemble fini de propositions comme l'a fait Endriss. À la place, les agents doivent donc construire leurs propositions. Pour cela, nous devons définir comment les agents distribuent l'utilité produite par les coalitions. Nous supposons que les agents souhaitent négocier sur le gain qu'ils recoivent à la formation des coalitions. Cependant, nous pouvons noter que lorsque les agents négocient ce gain au sein d'une coalition, la notion de structure n'a pas d'importance, et le cadre peut être ramené à une simple négociation multilatérale entre agents utilitaristes, c'est pourquoi le protocole d'Endriss peut-être utilisé pour cette négociation.

Dans le protocole d'Endriss, il est fait l'hypothèse que le gain minimal d'un agent est 0. Cependant, pour produire une solution rationnelle dans le cadre de la formation de coalitions, il faut que les coalitions proposées génèrent assez d'utilité pour chaque agent obtienne au moins l'utilité produite par sa coalition singleton. Nous faisons donc l'hypothèse que la valeur minimale pour chaque agent est le montant d'utilité produit par sa coalition singleton. Pour une coalition quelconque, la somme des valeurs singleton de

ses membres est alors réservée, afin que les agents puissent négocier sur le reste, appelé le *surplus*. Celui-ci décrit une notion différente du surplus défini dans la définition 3.12 du chapitre 3, cette dernière étant définie pour un agent individuel et dans un contexte stochastique.

Définition 5.8 (Surplus)**.** *Le* surplus S_C *d'une coalition* C *est :*

$$S_C = v(C) - \sum_{a \in C} v(a)$$

Il est intéressant de souligner le fait que nous pouvons utiliser une simple règle de répartition égalitaire de ce surplus au lieu d'exécuter le protocole, tout en obtenant le même résultat, comme l'indique le théorème suivant.

Theorème 5.1.1. *Soit une coalition* C *d'agents. L'application du protocole d'Endriss avec une stratégie PI et un type égocentrique pour la répartition des gains entre les membres de* C *est équivalent à attribuer à chaque agent une utilité égale à la valeur de leur coalition singleton plus une part égale du surplus de* C*.*

Démonstration. Soit une coalition C de n agents rationnels négociant leurs parts d'utilité au sein de cette coalition dans un cadre à utilité transférable. Sachant qu'un agent n'acceptera pas un gain inférieur à la valeur de sa coalition singleton, toute solution proposant un tel partage sera de fait rejetée. L'espace des propositions pouvant mener à un accord est donc composé des propositions qui assurent à chaque agent une utilité égale à celle de sa coalition singleton, plus une part de surplus. La négociation se fait donc sur le surplus uniquement. Soit la stratégie PI décrivant le fait que l'agent faisant la proposition dont le produit des utilités est le plus petit concédera.

Rappelons l'inégalité arthimético-géométrique qui est définie comme :

$$\frac{x_1 + \ldots + x_n}{n} \geq \sqrt[n]{x_1 \times \ldots \times x_n}$$

Il n'y a égalité entre les termes que si, et seulement si, $x_1 = x_2 = \ldots = x_n$. Ainsi, et en se rappelant que la fonction racine est une fonction strictement croissante pour tout nombre réel positif, nous pouvons déduire que le produit est maximal lors de l'égalité $x_1 = x_2 = \ldots = x_n$.

Or, si nous reprenons la coalition C et son surplus $S(C)$. Soit une répartition (efficace) de ce surplus réparti entre n agents : $\vec{x} = \{x_1, \ldots, x_n\}$, où x_i est la part de surplus de

l'agent a_i. Alors sa somme est telle que $\sum_{a_i \in C} x_i = S(C)$, et cette coalition possède un surplus moyen tel que $\frac{x_1+\ldots+x_n}{n}$. Pour une coalition donnée, cette moyenne arithmétique est constante (car $S(C)$ et n constants), et donc la répartition maximisant le produit est lorsque tous les agents gagnent la même part de surplus, c'est-à-dire $x_1 = x_2 = \ldots = x_n$. La valeur de Zeuthen de la stratégie PI atteint donc une valeur maximale (qui assure donc de ne pas concéder) avec une répartition égalitaire pour un surplus donné. L'espace des propositions étant infini dans un cadre à utilité transférable, les agents proposant toujours le gain maximal pour eux et concédant de manière égocentrique, la répartition vers laquelle on tend est donc la répartition totalement égalitaire du surplus, car celle-ci maximise le produit et la valeur de la stratégie. La répartition égalitaire du surplus entre les agents d'une coalition, en plus de l'utilité de leur coalition singleton, produit donc le même résultat que si le protocole d'Endriss avec la stratégie PI et le type égocentrique avait été appliqué. □

Ainsi, appliquer la stratégie PI d'un protocole de concessions monotones au sein de la coalition revient à répartir l'utilité de la coalition de sorte que chaque agent appartenant à celle-ci gagne le montant d'utilité de sa propre coalition singleton, auquel le surplus moyen (le surplus divisé par le nombre d'agents de la coalition) est additionné. Nous considérons donc une règle de distribution des gains rationnelle et égalitariste sur le surplus. Par définition, les coalitions singleton ont un surplus de 0. La règle est égalitaire car le surplus est alors distribué égalitairement entre les agents. Formellement,

Définition 5.9 (Part de surplus)**.** *La* part de surplus $S_C^{a_i}$ *d'un agent* a_i *dans sa coalition* C*, est le surplus de* C *divisé par son nombre d'agents :*

$$S_C^{a_i} = \frac{S_C}{|C|}$$

Définition 5.10 (Règle de distribution)**.** *Le gain de l'agent* a_i *appartenant à la coalition* C *avec un surplus* $S_C > 0$ *est défini comme :*

$$x_i^C = v(\{a_i\}) + S_C^{a_i}$$

Exemple 31. *Soit la fonction caractéristique v suivante :*

$$v = \Big\{\{\text{🔵}\} = 0.83\ ;\ \{\text{🔴}\} = 0.74\ ;\ \{\text{🟢}\} = 0.04\ ;$$
$$\{\text{🔵🔴}\} = 5.02\ ;\ \{\text{🔵🟢}\} = 0.81\ ;\ \{\text{🔴🟢}\} = 2.51\ ;$$
$$\{\text{🔵🔴🟢}\} = 2.65\ ;\ \{\emptyset\} = 0\Big\}$$

Une proposition satisfaisant notre règle de distribution pourrait être la suivante :

$$p_{\text{🔵}} = \Big(\Big[(\text{🔵}, \text{🔴}); (\text{🟢})\Big], (2.555, 2.465, 0.04)\Big)$$

Enfin, la notion d'accord commun s'en trouve donc modifiée, afin de correspondre à la nouvelle notion de proposition. Cependant, il faut souligner le fait que dû à notre règle de distribution, une même structure de coalitions ne correspond qu'à un seul vecteur de gain, tout le surplus d'utilité des coalitions étant réparti de façon égalitaire. La notion d'accord se porte donc naturellement sur la structure de coalitions des propositions. Pour rappel, dans le protocole originel, les agents n'ont pas besoin de tous proposer la même proposition pour qu'un accord soit atteint (voir définition 5.3). Nous faisons cependant l'hypothèse ici, dans le contexte de la formation de coalitions, qu'elles doivent l'être car les agents négocient une structure de coalitions.

Définition 5.11 (Accord commun)**.** *Un* accord commun *est atteint si, et seulement si, tous les agents font une proposition identique. Formellement :*

$$\forall\ a_i, a_j \in N \times N, p_i = p_j$$

Comme il a été dit précédemment, la notion de conflit n'a pas besoin d'être redéfinie.

5.1.3.2 Stratégies de concession

À chaque pas de temps, les agents font des propositions et, suivant une stratégie de concession, un ou plusieurs agents doivent abandonner leurs propositions et en faire de nouvelles. Pour cela, nous utilisons et adaptons la stratégie WRC présente dans les travaux d'Endriss. Nous n'adaptons pas la stratégie PI car comme il a été montré, cette stratégie favorise les distributions de gains égalitaires entre tous les agents, ce qui n'est pas adapté aux structures très diverses de la formation de coalitions. En effet, les gains de chaque agent dépendant de la coalition dans laquelle il se trouve et la stabilité dépendant des

coalitions pour lesquelles il pourrait dévier, cette notion de distribution égalitaire parmi des coalitions diverses. Nous proposons donc trois adaptations de la stratégie WRC que nous présentons ci-après : une stratégie est fondée sur les gains individuels des agents, une sur l'utilité des coalitions, et une dernière fondée sur le surplus des coalitions. Nous les appelons respectivement *WRC-classic*, *WRC-coalitions* et *WRC-surplus*.

Définition 5.12 (WRC-Classic). *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_i(p_i) - \min_{\forall j \in N} x_i(p_j)}{x_i(p_i)}$$

où $x_i(p_j)$ est le gain de l'agent a_i dans la proposition p_j.

Définition 5.13 (WRC-Coalitions). *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min_{\forall j \in N} x_{C_i}(p_j)}{x_{C_i}(p_i)}$$

où $x_{C_i}(p_j)$ est la somme des gains de tous les agents dans la coalition C_i dans la proposition p_j, où C_i est la coalition dans laquelle l'agent a_i est dans sa propre proposition :

$$x_{C_i}(p_j) = \sum_{k \in C_i} x_k(p_j)$$

Définition 5.14 (WRC-Surplus). *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{S_{C_i}(p_i) - \min_{\forall j \in N} S_{C_i}(p_j)}{S_{C_i}(p_i)}$$

où $S_{C_i}(p_j)$ est la somme des parts de surplus de tous les agents dans la coalition C_i dans la proposition p_j, où C_i est la coalition dans laquelle l'agent a_i est dans sa propre proposition :

$$S_{C_i}(p_j) = \sum_{k \in C_i} S^k_{C^j_k}$$

où C^j_k est la coalition de l'agent a_k dans la proposition p_j.

Dans toutes ces stratégies, la condition où $Z_{a_i} = 1$ (l'agent ne peut plus concéder) si l'agent a_i ne gagne rien dans sa propre proposition est adaptée pour que $Z_{a_i} = 1$ si,

et seulement si, l'agent a_i se propose lui-même dans sa coalition singleton ou dans toute coalition dont le surplus est nul. Une telle règle représente l'absence de coopération car aucun agent n'acceptera d'être irrationnel.

Enfin, quand un agent concède, il doit faire une nouvelle proposition selon un certain type. Comme indiqué précédemment, nous n'avons pas besoin d'adapter ces types à la formation de coalitions car ils s'appliquent parfaitement aux vecteurs de gains, sans nécessité de changement.

5.1.3.3 Déroulement du protocole

Les éléments du protocole étant désormais définis et adaptés, nous pouvons détailler les tours de négociation. Ici, l'ensemble des agents connaissent la fonction caractéristique complète, et chaque agent possède une liste noire privée servant à mémoriser les coalitions et structures de coalitions ayant été proposées puis rejetées, et celles qui ont été simplement rejetées car elles ne satisfaisaient pas le type de concession.

1. Chaque agent calcule le surplus des coalitions dont il peut faire partie,
2. Chaque agent crée une liste noire privée de coalitions et structures de coalitions, destinée à contenir les structures rejetées que ce soit par concession ou car ne respectant pas le type de concession. Les coalitions appartenant à des structures ayant toutes été ajoutées à la liste noire sont elles-mêmes ajoutées dans cette dernière.
3. Au premier tour, chaque agent fait une proposition initiale en choisissant sa coalition, notée C^*, parmi celles qui maximisent sa part de surplus, puis en choisissant la structure de coalitions qui maximise le bien-être social, notée $\mathcal{CS}^*$, et qui inclut la coalition choisie C^*,
4. À chaque tour suivant, chaque agent garde sa proposition ou concède, selon sa stratégie de concession,
5. Si un agent concède, la structure de coalitions précédente $\mathcal{CS}^*$ est ajoutée à la liste noire, et l'agent essaye de construire une nouvelle proposition satisfaisant son type de concession avec une autre structure de coalitions $\mathcal{CS}^{*'}$ qui inclut également sa coalition choisie C^*. Si une structure $\mathcal{CS}^{*'}$ ne satisfait pas un type de concession, elle est ajoutée à la liste noire. S'il n'y a plus de structure de coalitions possible avec la coalition C^*, cette dernière est ajoutée à la liste noire, et l'agent choisit une autre coalition, $C^{*'}$, qui maximise sa part de surplus. Un agent qui propose une

coalition $C^{*'}$ dont le surplus est égal à 0 se retire du processus car il ne pourra pas gagner quoique ce soit.

6. Répéter à partir de l'étape (4) jusqu'à ce qu'un accord soit atteint ou qu'aucun agent ne reste dans le processus, c'est-à-dire qu'aucun agent ne puisse faire une proposition où il gagnerait quelque chose (donc un *conflit*).

Si aucun agent ne peut faire une nouvelle proposition, le processus s'arrête et le protocole retourne la structure de coalitions singleton. C'est donc un protocole distribué car les agents exécutent certaines étapes localement (construction des propositions, gestion de leurs listes noires privées, calcul des concédants) tout en ayant une connaissance globale des propositions qui ont été faites et donc de celles qu'il reste à faire.

5.2 Un protocole de concession décentralisé

Afin de répondre à notre problématique de décentralisation, il nous faut compléter notre protocole de concessions pour la formation de coalitions présenté précédemment. En effet dans celui-ci, les agents possèdent des connaissances communes sur la fonction caractéristique, ce qui ne correspond pas à un cadre décentralisé, dans lequel les agents ne peuvent avoir de connaissances que sur les coalitions dont ils peuvent faire partie. Toutefois, relâcher cette hypothèse de connaissance commune a un impact important sur le protocole car les agents ne peuvent donc plus proposer de structures de coalitions mais simplement leur propre coalition, n'ayant de connaissances que pour choisir cette dernière et non une structure. Cependant, afin que les agents puissent construire leurs propositions, nous devons alors faire l'hypothèse qu'ils se communiquent les utilités de leurs coalitions singleton. Ces communications sont supposées parfaites et correctes. En effet, sans cette dernière connaissance commune, le protocole ne pourra pas être décentralisé, car sans cela, les agents ne pourront pas utiliser efficacement la règle de distribution. Il est à noter que cette hypothèse n'est pas une hypothèse forte car nous pouvons simplement étendre le protocole en faisant un tour d'initialisation où chaque agent annonce la valeur de sa coalition singleton.

Enfin, nous devons également redéfinir les stratégies de concession, car outre le fait que nous revenons à des propositions sans stratégies, les agents ne proposent plus un vecteur de gains concernant tous les agents, mais uniquement ceux de leur coalition. Il faut donc que les agents puissent raisonner sur ces vecteurs restreints.

5.2.1 Adaptation des notions à la décentralisation

Afin de décentraliser le protocole en s'abstrayant de la connaissance commune des agents, nous devons redéfinir la fonction caractéristique de sorte à ce que les agents ne connaissent que les utilités qui les concernent, et donc les coalitions dont ils font partie. Cependant, cela impacte le protocole qui ne peut donc plus négocier sur les structures de coalitions, mais uniquement sur les coalitions. La redéfinition de la fonction caractéristique implique donc une redéfinition des propositions.

5.2.1.1 Fonction caractéristique

Afin de réduire les connaissances des agents, nous redéfinissons simplement la fonction caractéristique. En effet, les agents ne doivent plus posséder une connaissance commune de celle-ci, c'est pourquoi chaque agent devra posséder une fonction caractéristique individuelle, ne portant que sur les coalitions dont il fait partie.

Définition 5.15 (Fonction caractéristique individuelle)**.** *Soit v_i la* fonction caractéristique individuelle *de l'agent a_i. Celle-ci est définie comme étant la restriction de la fonction caractéristique v d'un jeu $\mathcal{G}$ à G_i l'ensemble de coalitions auxquelles a_i appartient :*

$$G_i = \{C \mid \forall C \in 2^N \ t.q.\ a_i \in C\}$$

La fonction v_i est donc définie comme suit :

$$v_i : G_i \to \mathbb{R}$$

Il est important de noter que la fonction caractéristique v est inconnue des agents.

Exemple 32. *Soit la fonction caractéristique v définie dans l'exemple 31 (page 116). La fonction caractéristique individuelle de*  *est :*

$$v_{\text{  }} = \Big\{\{\text{  }\} = 0.83\ ;\ \{\text{  }\} = 0.81\ ;\ \{\text{  }\} = 5.02\ ;\ \{\text{  }\} = 2.65\Big\}$$

5.2.1.2 Propositions

Les connaissances des agents étant plus restreintes, ces derniers ne peuvent plus proposer de structures de coalitions. C'est pourquoi les propositions doivent être redéfinies pour le cadre décentralisé afin que les agents ne proposent plus qu'une coalition et les

gains pour les agents appartenant à cette dernière. Il faut cependant rappeler que nous faisons l'hypothèse qu'avant que les agents construisent leurs propositions, ils échangent la connaissance sur les utilités de leurs coalitions singletons.

Définition 5.16 (Proposition décentralisée)**.** *Étant donné un jeu $\mathcal{G}$, une* proposition *dans un cadre* décentralisé *de l'agent a_i, notée p_i, est un tuple $S^{dec}_{\mathcal{G}} = \langle C, \vec{x} \rangle$ où $\vec{x}$ est un vecteur de gains $\langle x_j^{C_j^i} \rangle$ où C_j^i est la coalition de l'agent a_j dans la proposition p_i.*

Exemple 33. *Soit la fonction caractéristique v définie dans l'exemple 31 (page 116). Un exemple de proposition décentralisée est :*

$$p_{\text{  }} = \left(\left[(\text{  }, \text{  }) \right], (2.555, 2.465) \right)$$

5.2.1.3 Définition d'un accord

Les agents ne proposant plus des structures de coalitions mais uniquement des coalitions, la notion d'accord doit être elle-aussi redéfinie. Il va sans dire que hormis le cas où les agents proposent tous la grande coalition, il est impossible que les propositions soient toutes les mêmes. C'est pourquoi nous introduisons la notion d'*accord local*.

Définition 5.17 (Accord local)**.** *Un* accord local *pour un ensemble d'agent Z existe si, et seulement si, pour tout agent $a_i \in Z$, $C_i^i = Z$,* et *que aucun agent $a_j \notin Z$ ne propose de coalition C_k^j telle que $a_k \in Z \ \forall \ a_k \in N$.*

Autrement dit, aucun agent appartenant à la coalition C ne doit être proposé dans une autre coalition que cette dernière. Le fait qu'un accord local sur une coalition C ne puisse pas exister lorsqu'un agent extérieur à C fait une proposition contenant un des agents de C se justifie par le fait qu'il faut que les agents de C soient certains qu'ils ne devraient pas concéder, i.e. que personne d'autres leur propose autre chose.

5.2.2 Rédéfinition du protocole

Comme les accords et les propositions ont été redéfinies, nous devons également redéfinir les stratégies de concessions, mais également les tours de négociation.

5.2.2.1 Stratégies de concession

La seule différence entre les versions distribuées et décentralisées des stratégies que nous proposons réside dans l'ensemble des propositions considérées par la stratégie. Nous

notons cet ensemble **E** et il décrit quelles propositions seront considérées par la stratégie afin de calculer la valeur de concession. Cette notation nous permet de mettre en lumière le passage d'un contexte distribué à un contexte décentralisé : les agents ayant des connaissances restreintes, il ne peuvent pas raisonner sur l'ensemble complet des propositions, mais uniquement celles d'intérêt pour eux.

Plus formellement, dans le cadre distribué, cet ensemble **E** est équivalent à l'ensemble **P** des propositions des agents, tel que $\mathbf{P} = \{p_j \mid \forall a_j \in N\}$. Dans les versions décentralisées, cet ensemble **P** est composé des agents qui proposent a_i dans leurs propres propositions, à l'exception des agents faisant la même proposition que a_i. Ce choix est justifié par le fait que les propositions faites dans le cadre décentralisé sont uniquement des coalitions, et donc une solution partielle qui ne contient pas tous les agents (hormis la proposition de la grande coalition). De ce fait, lorsqu'un agent n'est proposé que par des agents faisant la même proposition que lui, s'il n'y a pas d'accord local en raison d'un agent tierce (voir définition 5.17), alors l'agent aura une valeur Z de 0, et donc devra nécessairement concéder.

Exemple 34. *Reprenons la fonction caractéristique v définie dans l'exemple 31 (page 116). Soient les propositions suivantes :*

$$p_{\bullet} = \Big(\big[(\bullet,\bullet)\big], (2.555, 2.465)\Big) \; ; \; p_{\bullet} = \Big(\big[(\bullet,\bullet)\big], (2.555, 2.465)\Big) \; ;$$
$$p_{\bullet} = \Big(\big[(\bullet,\bullet)\big], (1.605, 0.905)\Big)$$

Si nous ne restreignons pas l'ensemble $\boldsymbol{P}_{\bullet}$ aux propositions différentes de $p_{\bullet}$, alors dans un tel exemple, l'agent $\bullet$ aura une valeur Z égale à 0 et devra forcément concéder, qu'importe sa proposition et celle des autres.

L'ensemble $\mathbf{P}_i$ pour un agent a_i ne contient donc que les agents a_j qui proposent a_i tels que $p_j \neq p_i$.

Cependant, il est alors possible que $\mathbf{P}_i$ soit vide. Afin de garder la sémantique propre à la stratégie WRC originelle qui est que l'agent qui concède est celui qui y perd le moins, et en l'absence de propositions complètes, nous faisons l'hypothèse que les agents pour lesquels l'ensemble $\mathbf{P}_i$ est vide considèrent la proposition suivante qu'ils feraient en cas de concession. Cette proposition future est donc simplement la deuxième coalition maximisant le surplus de l'agent après la coalition qu'il est en train de proposer. Ainsi, si aucun agent ne le propose, l'agent peut évaluer ses pertes en cas de concession. Ainsi, la

définition de l'ensemble $\mathbf{E}_i$ pour un agent a_i est telle que suit.

Définition 5.18 (Ensemble restreint de propositions)**.** *Un* ensemble restreint de propositions *pour un agent a_i, noté $\boldsymbol{E}_i$, est l'ensemble des propositions que l'agent a_i va considérer dans une stratégie. Cet ensemble est équivalent à l'ensemble $\boldsymbol{P}_i = \{p_j \mid \forall a_j \in N, a_i \in p_j \wedge p_j \neq p_i\}$ des propositions p_j différentes de p_i telles que $a_i \in p_j$ si celui-ci est non-vide. Sinon, l'ensemble $\boldsymbol{E}_i$ ne contiendra que la proposition suivante de l'agent a_i, notée $p_i^{\rightarrow}$.*

$$\boldsymbol{E}_i = \begin{cases} \boldsymbol{P}_i & si \ \ |\boldsymbol{P}_i| > 0, \\ \{p_i^{\rightarrow}\} & sinon. \end{cases}$$

La proposition suivante $p_i^{\rightarrow}$ de l'agent a_i est simplement la coalition (et sa distribution de gains) qui maximise le surplus de l'agent a_i après la coalition proposée dans p_i, et qui respecte le type de concession.

Exemple 35. *Reprenons la fonction caractéristique v définie dans l'exemple 31 (page 116). Soient les propositions suivantes :*

$$p_{\color{blue}\bullet\color{red}\bullet} = \left(\left[(\color{blue}\bullet\color{black}, \color{red}\bullet\color{black})\right], (2.555, 2.465)\right) \ ; \ p_{\color{red}\bullet} = \left(\left[(\color{blue}\bullet\color{black}, \color{red}\bullet\color{black})\right], (2.555, 2.465)\right) \ ;$$
$$p_{\color{green}\bullet} = \left(\left[(\color{red}\bullet\color{black}, \color{green}\bullet\color{black})\right], (1.605, 0.905)\right)$$

Les propositions suivantes que les agents feraient en cas de concessions sont les suivantes.

$$p^{\rightarrow}_{\color{blue}\bullet} = \left(\left[(\color{blue}\bullet\color{black}, \color{red}\bullet\color{black}, \color{green}\bullet\color{black})\right], (1.177, 1.087, 0.387)\right) \ ; \ p^{\rightarrow}_{\color{red}\bullet} = \left(\left[(\color{red}\bullet\color{black}, \color{green}\bullet\color{black})\right], (1.605, 0.905)\right) \ ;$$
$$p^{\rightarrow}_{\color{green}\bullet} = \left(\left[(\color{blue}\bullet\color{black}, \color{red}\bullet\color{black}, \color{green}\bullet\color{black})\right], (1.177, 1.087, 0.387)\right)$$

Les ensembles restreints de propositions des agents sont comme suit. $\boldsymbol{E}_{\color{blue}\bullet} = \{ p^{\rightarrow}_{\color{blue}\bullet} \}$ car bien que l'agent soit proposé par un autre, cela est la même proposition, elle n'est donc pas considérée dans l'ensemble $\boldsymbol{E}$, sa proposition future est donc ajoutée à l'ensemble. $\boldsymbol{E}_{\color{red}\bullet} = \{ p_{\color{red}\bullet} \}$ car l'agent est proposé par un autre. $\boldsymbol{E}_{\color{green}\bullet} = \{ p^{\rightarrow}_{\color{green}\bullet} \}$ car l'agent n'est proposé par aucun autre, sa proposition future est donc ajoutée à son ensemble.

Nous pouvons donc désormais redéfinir les stratégies distribuées en s'appuyant de cet ensemble.

Définition 5.19 (WRC-Classic-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur*

Z_{a_i} est la plus petite, où :

$$Z_{a_i} = \frac{x_i(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_i(p_j)}{x_i(p_i)}$$

où $x_i(p_j)$ est le gain de l'agent a_i dans la proposition p_j.

Tandis que la stratégie WRC-Classic est adaptable simplement, ce n'est pas le cas pour WRC-Coalitions, en raison de l'absence d'informations sur les gains des autres agents dans les diverses propositions. C'est pourquoi lorsqu'un agent n'appartient à une proposition, nous considérons sa valeur singleton dans le calcul de la valeur de la coalition, comme nous le permet l'hypothèse que nous avons faite concernant la communication des utilités des coalitions singletons.

Définition 5.20 (WRC-Coalitions-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_{C_i}(p_j)}{x_{C_i}(p_i)}$$

où $x_{C_i}(p_j)$ est la somme des gains de tous les agents dans la coalition C_i dans la proposition p_j, où C_i est la coalition dans laquelle l'agent a_i est dans sa propre proposition :

$$x_{C_i}(p_j) = \sum_{k \in C_i \cap C_j} x_k(p_j) + \sum_{k \in C_i, k \notin C_j} v(\{a_k\})$$

Concernant les types de concessions, comme pour le cadre distribué, aucune adaptation n'est nécessaire, hormis le fait que seuls les agents qui apparaîssent dans les deux propositions sont pris en compte par le type de concession, en raison de l'information manquante dans le cas contraire.

5.2.2.2 Étapes du protocole

Le protocole doit donc être adapté aux nouvelles définitions des propositions et d'un accord. Pour cela, toutes les parties concernant les structures de coalitions sont supprimées, et des étapes sont rajoutées pour l'accord : lorsqu'un (ou plusieurs) *accord local* est trouvé, les agents formant cet accord se retirent du processus, alors les agents recommencent un tour complet. De plus, le cas où aucun agent ne reste dans le processus est affiné pour prendre en compte l'éventualité où tous les agents ont trouvé un accord local.

Ici, les agents ne connaissent que leur propre fonction caractéristique individuelle, ainsi qu'une liste noire privée servant à mémoriser les coalitions ayant été proposées et rejetées, et celles rejetées en raison du type de concession. De plus, il est fait l'hypothèse qu'au début du protocole, les agents s'échangent la connaissance de l'utilité de leurs coalitions singleton afin de pouvoir construire leurs propositions.

1. Chaque agent calcule le surplus des coalitions dont il peut faire partie,
2. Chaque agent crée une liste noire privée de coalitions,
3. Au premier tour, chaque agent fait une proposition initiale en choisissant sa coalition, notée C^*, parmi celles qui maximisent sa part de surplus,
4. À chaque tour suivant, chaque agent garde sa proposition ou concède, selon sa stratégie de concession, sauf si un *accord local* existe, alors les agents concernés se retirent du processus, les autres agents répètent le processus à partir de l'étape (3),
5. Si un agent concède, la coalition C^* est ajoutée à la liste noire, et l'agent choisit une autre coalition, $C^{*'}$, qui maximise sa part de surplus. Un agent qui propose une coalition $C^{*'}$ dont le surplus est égal à 0 se retire du processus car il ne pourra pas gagner quoique ce soit.
6. Répéter à partir de l'étape (4) jusqu'à ce que tous les agents possèdent un accord local (ce qui correspond donc à un accord total) ou qu'aucun agent ne reste dans le processus, c'est-à-dire qu'aucun agent ne puisse faire une proposition où il gagnerait quelque chose (donc un *conflit*).

La différence entre ce protocole décentralisé et le protocole distribué de la section précédente réside donc dans le fait que les propositions ne concernent plus nécessairement l'ensemble des agents, c'est pourquoi l'accord local a été défini. Afin d'empêcher tout biais dû à cet accord local, un tour complet de négociation est rejoué lorsque l'accord local survient. Tout comme pour le protocole distribué, si les agents (hormis ceux qui ont un accord local) ne peuvent plus faire de nouvelle proposition, le processus s'arrête. Toutefois, si des accords locaux ont été trouvés par certains agents, les accords seront respectés et les coalitions correspondantes formées, tandis que le reste des agents formeront leurs coalitions singleton.

5.3 Expérimentations

Pour évaluer l'intérêt de notre protocole, nous procédons empiriquement. En particulier, un des objectifs est de comparer l'influence des différents types et stratégies de concession sur notre protocole. Pour cela, nous générons des jeux aléatoires avec différentes fonctions caractéristiques et appliquons le protocole plusieurs fois sur chaque jeu, avec différents paramètres pour les agents. Ces paramètres sont le type et la stratégie de concession utilisés.

5.3.1 Paramètres globaux des expérimentations

Les expérimentations menées possèdent des paramètres communs afin de pouvoir comparer les protocoles dans différents cadres, c'est notamment le cas des jeux utilisés, qui seront communs à toutes les expérimentations, mais aussi du nombre d'agents.

5.3.1.1 Classe NDCS

Suite aux expérimentations précédentes, nous avons décidé de nous concentrer sur une seule classe de fonctions caractéristiques, et celle-ci est la classe NDCS (voir section 4.1.1.1). Ainsi, chaque coalition $C \subseteq N$ possède une utilité $v(C)$ donnée par une variable aléatoire $\mathcal{N}(|C|, \sqrt{|C|})$, avant d'être normalisée sur l'invervalle $[0, 1]$.

5.3.1.2 Construction des jeux et paramètres

Nous construisons 100 jeux avec 8 agents, chaque jeu ayant une fonction caractéristique différente. Nous faisons l'hypothèse que les agents sont homogènes, c'est-à-dire qu'ils utilisent tous la même stratégie et le même type de concession. S'il est nécessaire que les agents aient la même stratégie de concession afin qu'ils soient tous d'accord sur qui concède, ce n'est pas le cas pour le type de concession. Cependant nous nous contentons de cette homogénéité pour cette première étude.

Notons que le nombre limité d'agents que nous utilisons n'est pas lié à la complexité du protocole mais au calcul combinatoire de la solution optimale pour le jeu afin de comparer notre protocole à celle-ci. En effet, nous désirons comparer notre solution à la solution optimale pour laquelle il n'existe pas d'algorithme de calcul efficace.

Les protocoles sont exécutés sur chacun des jeux avec chaque couple de stratégie et type de concession possible.

5.3.2 Métriques

Nous considérons quatre mesures empiriques, respectivement fondées sur le dernier cœur, sur le meilleur ϵ-cœur atteignable par notre protocole, le bien-être social, et le nombre de Bell, c'est-à-dire le nombre de partitions possibles pour un nombre d'agents donné [Bell, 1938, Rota, 1964].

Les deux premières mesurent la distance entre les solutions trouvées par le protocole par rapport au dernier cœur, ce dernier comprenant les meilleures solutions stables que l'on puisse trouver. Ces mesures nous permettent d'évaluer la perte de stabilité due à la négociation et à notre règle de distribution spécifique du surplus.

La troisième est le *prix de la stabilité* [Anshelevich *et al.*, 2008] (voir section 1.3.1.3), qui, pour rappel, mesure le gain que les agents doivent abandonner, par rapport au bien-être social maximal, afin de former une structure stable.

La dernière mesure correspond au nombre de structures de coalitions qui ont été explorées, c'est-à-dire le nombre de structures qui ont été sélectionnées à un moment dans le protocole, sans être nécessairement proposées. Afin d'avoir des données commensurables et comparables, chaque mesure est définie comme un ratio, avec des valeurs comprises sur l'intervalle $[0, 1]$.

5.3.2.1 Distance au dernier cœur

La première mesure est le ratio entre la valeur ϵ de l'ϵ-cœur auquel appartient la solution trouvée par le protocole et la valeur ϵ du dernier cœur, c'est-à-dire le minimum atteignable. Ainsi, si le dernier cœur n'est pas un 0-cœur, les solutions du protocole ne seront pas désavantagées contrairement à une comparaison directe au 0-cœur.

Définition 5.21 (Ratio au dernier cœur)**.** *Étant donné un jeu $\mathcal{G}$, soient $\epsilon^*(\mathcal{G})$ la valeur ϵ du dernier cœur, et $\epsilon^p(\mathcal{G})$ la valeur ϵ de l'ϵ-cœur auquel appartient la solution retournée par le protocole. Le ratio au dernier cœur, noté $R^*(\mathcal{G})$, est défini par :*

$$R^*(\mathcal{G}) = \frac{1 - \epsilon^p(\mathcal{G})}{1 - \epsilon^*(\mathcal{G})}$$

5.3.2.2 Distance à l'optimal-protocole

La seconde mesure est le ratio entre la valeur ϵ de l'ϵ-cœur auquel appartient la solution trouvée par le protocole et la valeur ϵ du meilleur ϵ-cœur (c'est-à-dire ayant la valeur ϵ

la plus faible) atteignable par notre protocole s'il couvre toutes les partitions possibles, selon notre règle de distribution.

Définition 5.22 (Ratio à l'optimal-protocole)**.** *Étant donné un jeu $\mathcal{G}$, soient $\epsilon^{p^*}(\mathcal{G})$ la valeur ϵ de l'ϵ-cœur optimal selon notre protocole et la règle de distribution, et $\epsilon^{p}(\mathcal{G})$ la valeur ϵ de l'ϵ-cœur auquel appartient la solution retournée par le protocole. Le ratio à l'optimal-protocole, noté $R^p(\mathcal{G})$, est défini par :*

$$R^p(\mathcal{G}) = \frac{1 - \epsilon^p(\mathcal{G})}{1 - \epsilon^{p^*}(\mathcal{G})}$$

5.3.2.3 Prix de la stabilité

La troisième mesure, définie par Anshelevich *et al.* [Anshelevich *et al.*, 2008], est le ratio entre le bien-être social de la solution trouvée par le protocole, qui est stable, et le bien-être social maximal pour le jeu.

Définition 5.23 (Prix de la stabilité)**.** *Étant donné un jeu $\mathcal{G}$, soient $\Pi(\mathcal{G})$ l'ensemble de toutes les structures de coalitions pour $\mathcal{G}$, et $S^p_{\mathcal{G}} = \langle \mathcal{CS}^p, \vec{x}^p \rangle$ une solution au jeu $\mathcal{G}$ retournée par le protocole. Le* prix de la stabilité *pour cette solution, noté $PS(S^p_{\mathcal{G}})$, est défini par :*

$$PS(S^p_{\mathcal{G}}) = \frac{\sum_{C \in \mathcal{CS}^p} v(C)}{\max_{\pi \in \Pi(\mathcal{G})} \sum_{C' \in \pi} v(C')}$$

5.3.2.4 Ratio de Bell

La dernière mesure est la proportion de structures de coalitions explorées par le protocole, c'est-à-dire qui ont été proposées ou sélectionnées puis mises dans la liste noire durant le protocole, comparé au nombre total de structures possibles, donné par le nombre de Bell. Étant donné que chaque agent peut proposer chaque partition une seule fois, et que leur liste noire est individuelle, le treillis des structures de coalitions peut être exploré autant de fois qu'il y a d'agents.

Définition 5.24 (Ratio de Bell)**.** *Étant donné le nombre de partitions explorées par l'agent a_i et le nombre de Bell, respectivement notés p^i_e et B_n (avec $n = |N|$), le ratio de Bell, noté $B^{\%}$, est :*

$$B^{\%} = \frac{\sum_{i \in N} p^i_e}{n \times B_n} \text{ avec } B_{m+1} = \sum_{k=0}^{m} \binom{m}{k} B_k$$

5.4 Résultats : distribué contre centralisé

Nous pouvons désormais exécuter le protocole dans un cadre distribué, et comparer son efficacité à une résolution centralisée grâce aux mesures proposées. Tout d'abord, fixons les paramètres spécifiques à ces expérimentations.

5.4.1 Paramètres des expérimentations

Afin d'évaluer l'efficacité du protocole distribué aux solutions calculées de manière centralisée, mais aussi pour faire une comparaison entre les types de concession, nous procédons empiriquement, à l'aide des quatres mesures définies précédemment, qui sont les ratios au dernier cœur, à l'optimal-protocole et de Bell, ainsi que le prix de la stabilité. Les différents jeux décrits en section 5.3.1 sont donc exécutés plusieurs fois avec des paramètres différents pour les agents, à savoir le type et la stratégie de concession utilisés. Tous les types sont utilisés, chacun avec chaque stratégie suivante : *WRC-Classic*, *WRC-Coalitions* et *WRC-Surplus*.

5.4.2 Lecture des graphiques

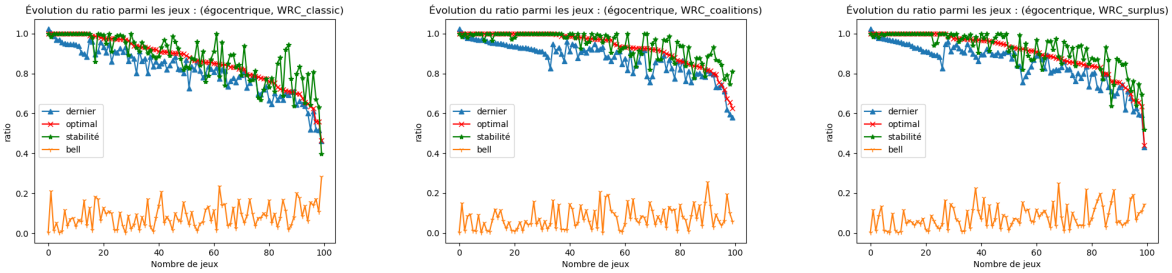
Les figures 5.4 et 5.5 montrent le ratio au dernier cœur, le ratio à l'optimal-protocole, le prix de la stabilité et le ratio de Bell pour chaque stratégie de concession possible. La figure 5.4 montre les résultats pour les types de concession *faible*, *égocentrique*, *Nash* et *utilitaire*, et la figure 5.5 les résultats pour les types *fort*, *égalitaire* et *Pareto*. Les données sont triées sur le ratio de l'optimal-protocole du plus grand au plus petit, et les trois autres mesures suivent ce tri pour rester liées au jeu correspondant. Également, les figures sont présentées dans un ordre particulier qui permet au lecteur de constater une évolution entre les résultats en fonction des différents types de concession, de ce qui semble empiriquement le moins bon, vers le meilleur. Concernant les différentes mesures et leurs significations graphiques : plus le ratio de Bell, étiquetté `bell`, est proche de 0, moins il y a eu de partitions explorées. Pour les trois autres mesures, la meilleure valeur est 1. Pour le ratio au dernier cœur, étiquetté `dernier`, et le ratio à l'optimal-protocole, étiquetté `optimal`, cela signifie que les solutions retournées par le protocole sont respectivement présentes dans le dernier cœur du jeu, et présentes dans le dernier cœur du jeu connaissant la règle de distribution. Pour le prix de la stabilité, étiquetté `stabilité`, cela signifie que les solutions retournées par le protocole maximisent le bien-être social. Si ces mesures

s'éloignent de 1, cela montre une perte d'efficacité.

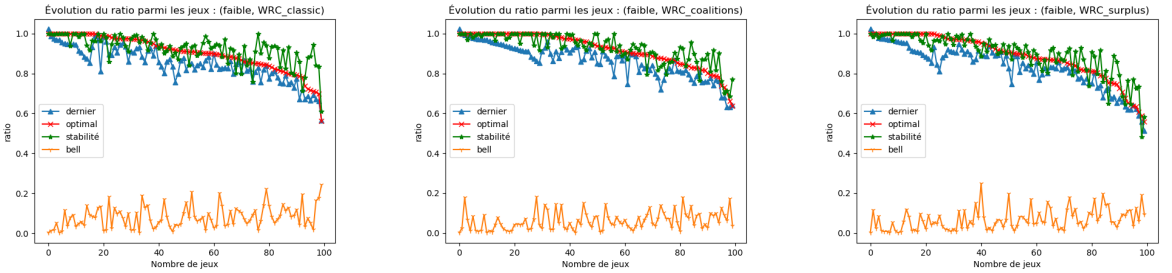
La figure 5.3 compare le ratio à l'optimal-protocole des trois stratégies sur trois différents types de concession : *faible*, *Nash* et *fort*.

5.4.3 Analyse des résultats

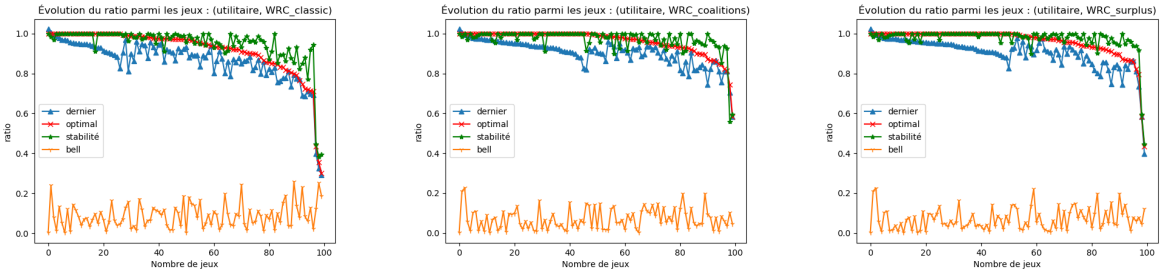
Tout d'abord, nous pouvons voir que la courbe du ratio de Bell a une forme similaire pour toutes les paires (type, stratégie). Nous pouvons estimer l'exploration moyenne sur tous les jeux à 10%, indépendemment du type et de la stratégie de concession. De plus, nous pouvons voir qu'il n'y a pas de corrélation apparente entre le fait d'explorer et une meilleure stabilité, ni l'inverse. Concernant le ratio à l'optimal-protocole, nous pouvons voir que les stratégies et les types de concession influent sur les résultats. Avec les types de concession *égocentrique*, *faible*, *utilitaire* et *Nash*, le ratio à l'optimal-protocole atteint des pertes de 40%, avec parfois des effrondements atteignant les 60% pour certaines paires de paramètres. Bien qu'elle comprenne les plus gros effondrements, la stratégie WRC-Surplus comprend avec la stratégie WRC-Coalitions le meilleur taux de valeurs optimales. Plus précisément, sur les trois derniers types de concession (*fort*, *égalitaire* et *Pareto*), la stratégie WRC-Classic atteint des pertes maximales d'environ 30%, contre 20% pour les deux autres stratégies. Ces deux stratégies ont donc les meilleurs résultats, avec environ 60% des jeux avec un ratio à l'optimal-protocole de 1. Les deux dernières mesures, le ratio au dernier cœur et le prix de la stabilité, se comportent de la même manière, à savoir qu'elles comportent des pics d'effondrement sur les mêmes jeux. Pour les types *faible* et *égocentrique*, nous pouvons voir de grandes pertes d'efficacité (régulièrement au dessus de 30%), et ce pour toutes les stratégies. En s'intéressant aux différences entre les stratégies sur ces types, nous pouvons voir que les pertes sont plus limitées avec WRC-Coalitions. Les valeurs entre WRC-Classic et WRC-Surplus sont proches, bien que cette dernière semble décroître un peu moins. Pour les types *Nash* et *utilitaire*, les pertes sont moins grandes. Elles sont de l'ordre de 20%, bien qu'il y ait des effondrements autour de 60%, comme avec le ratio à l'optimal-protocole. Pour le type *Nash*, les pertes de bien-être social sont réduites avec WRC-Classic, mais WRC-Coalitions est légèrement meilleure pour le ratio au dernier cœur. Pour le type *utilitaire*, ce sont WRC-Coalitions et WRC-Surplus qui minimisent les pertes sur les mesures. Enfin, pour les trois derniers types (*fort*, *égalitaire* et *Pareto*), elles semblent toutes similaires. Il y a en réalité des différences mineures sur le nombre de partitions explorées et le nombre de concessions effectuées, mais les résultats du protocole sont identiques. Ce phénomène vient, à notre avis, de



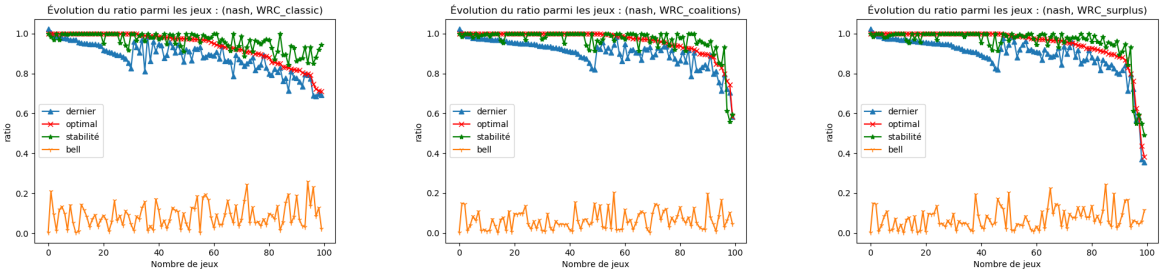
(a) Égo. / WRC-Classic (b) Égo. / WRC-Coalitions (c) Égo. / WRC-Surplus



(d) Faible / WRC-Classic (e) Faible / WRC-Coalitions (f) Faible / WRC-Surplus

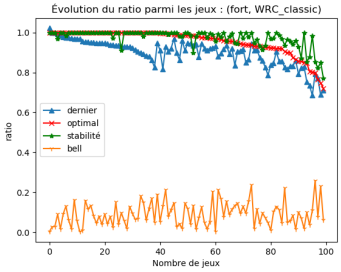


(g) Util. / WRC-Classic (h) Util. / WRC-Coalitions (i) Util. / WRC-Surplus

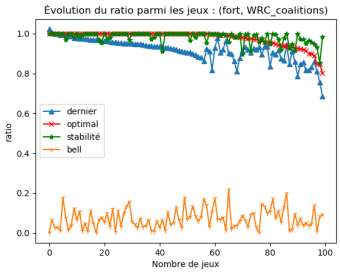


(j) Nash / WRC-Classic (k) Nash / WRC-Coalitions (l) Nash / WRC-Surplus

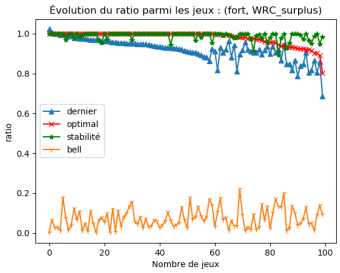
FIGURE 5.1 – Résultats pour un couple (Type / Stratégie) de concession



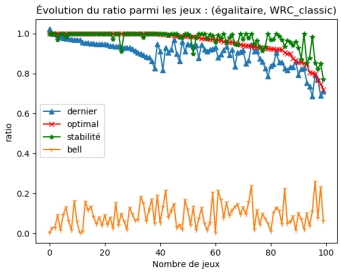
(a) Fort / WRC-Classic



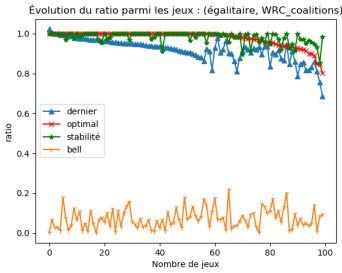
(b) Fort / WRC-Coalitions



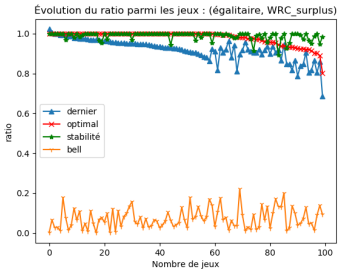
(c) Fort / WRC-Surplus



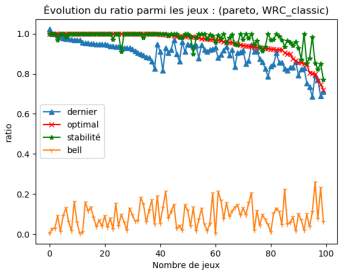
(d) Éga. / WRC-Classic



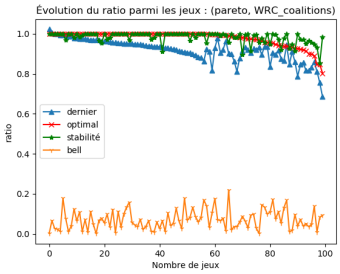
(e) Éga. / WRC-Coalitions



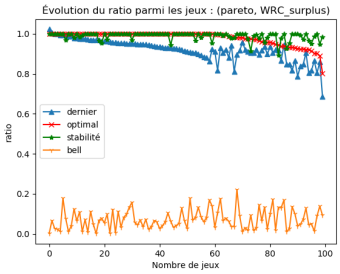
(f) Éga. / WRC-Surplus



(g) Pareto / WRC-Classic

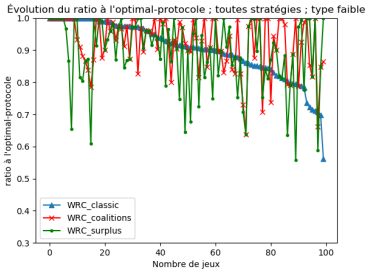


(h) Pareto / WRC-Coalitions

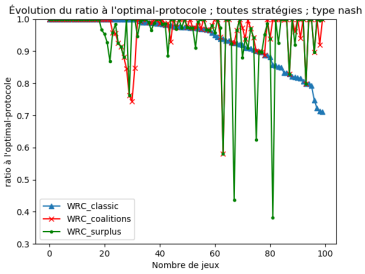


(i) Pareto / WRC-Surplus

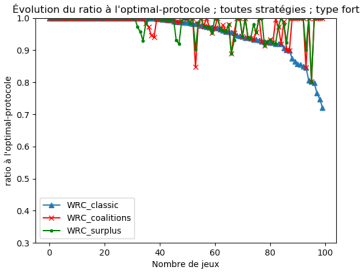
FIGURE 5.2 – Résultats pour un couple (Type / Stratégie) de concession



(a) Faible



(b) Nash



(c) Fort

FIGURE 5.3 – Exemples de différences entre stratégies

la règle de distribution choisie, qui doit contraindre davantage ces types de concession et les rapproche donc sémantiquement. Toutefois, il doit être noté que ces trois types sont empiriquement meilleurs que les autres, et ce notamment avec les stratégies WRC-Coalitions et WRC-Surplus, où les pertes sur le ratio au dernier cœur atteignent moins souvent 20%, et où le prix de la stabilité excède très rarement 10% tout en étant souvent à 0%, c'est-à-dire optimal. Concernant la figure 5.3, nous choisissons de montrer seulement ces trois types car comme montré sur les figures précédentes, la tendance des courbes est semblable pour certains types : *faible* et *égocentrique*, *Nash* et *utilitaire*, *égalitaire*, *fort* et *Pareto*. Les données affichées sont triées de la plus grande à la plus petite valeur de ratio à l'optimal-protocole avec la stratégie WRC-Classic, et les valeurs pour les autres stratégies suivent ce tri en restant liées au jeu correspondant. La figure 5.3.a montre des valeurs qui varient beaucoup, mais celles-ci décroissent moins et moins souvent pour WRC-Coalitions. WRC-Surplus semble être la moins bonne stratégie de ce cas. Concernant le type *Nash*, WRC-Surplus semble également la pire stratégie, avec des pertes atteignant 60%. Comme précédemment, WRC-Coalitions semble être la meilleure. Pour le type *fort*, les valeurs s'effondrent beaucoup moins, avec une perte maximale de 30% environ. Ici, WRC-Coalitions et WRC-Surplus semblent proches, et presque toujours plus efficaces que WRC-Classic.

Pour résumer, trois types de concession minimisent les pertes plus que les autres : *fort*, *égalitaire* et *Pareto*. Pour les stratégies, WRC-Coalitions semble être celle avec les meilleurs résultats, suivie de près par WRC-Surplus (qui manque cependant d'efficacité sur certains types de concession). Ces résultats sont cohérents : une perte d'optimalité due à la distribution et des stratégies adaptées plus efficaces.

5.5 Résultats : décentralisé contre distribué

Nous pouvons désormais exécuter le protocole dans un cadre décentralisé, et comparer son efficacité au protocole exécuté dans un cadre distribué. Les résultats pour la résolution centralisée et pour le protocole distribué sont donc les mêmes qu'en section 5.4. Fixons les paramètres spécifiques à ces nouvelles expérimentations.

5.5.1 Paramètres des expérimentations

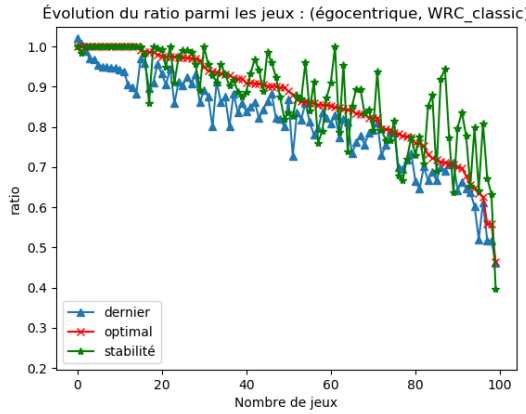
Afin d'évaluer l'efficacité du protocole décentralisé comparé au protocole distribué et à une solution calculée de manière centralisée, mais aussi pour faire une comparaison entre les stratégies décentralisées et les types de concession, nous procédons à nouveau empiriquement, avec trois des quatres mesures établies précédemment. En effet, le ratio de Bell n'est pas réutilisé car les agents n'explorent pas les mêmes ensembles dans les différents protocoles, à savoir l'ensemble des structures de coalitions pour la version distribuée, et l'ensemble des coalitions pour la version décentralisée. Nous considérons donc trois mesures empiriques qui sont les ratios au dernier cœur et à l'optimal-protocole et le prix de la stabilité. Les différents jeux décrits en section 5.3.1 sont donc exécutés plusieurs fois avec des paramètres différents pour les agents, à savoir le type et la stratégie de concession utilisés. Les types restent inchangés et au nombre de sept, tandis que les stratégies sont *WRC-Classic-Dec* et *WRC-Coalitions-Dec*.

5.5.2 Lecture des graphiques

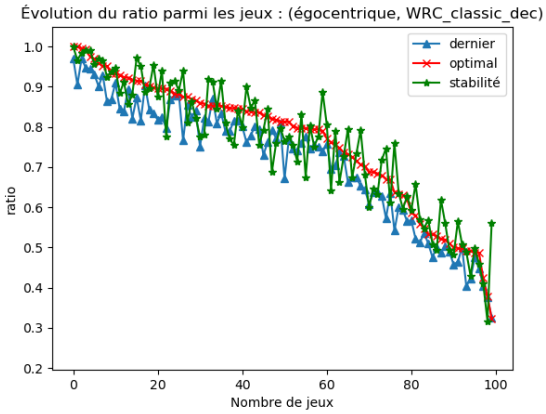
Les figures 5.4, 5.5 et 5.6 montrent les ratios au dernier cœur et à l'optimal-protocole ainsi que le prix de la stabilité pour les stratégies de concessions WRC-Classic et WRC-Classic-Dec. Les figures 5.7, 5.8 et 5.9 montrent ces mêmes mesures pour les stratégies WRC-Coalitions et WRC-Coalitions-Dec. Les données sont triées du ratio à l'optimal-protocole le plus grand au plus petit, et les deux autres mesures suivent ce tri pour rester liées au jeu correspondant, celles-ci étant bien représentées par le ratio à l'optimal-protocole, sur lequel nous allons cibler notre analyse.

5.5.3 Analyse des résultats

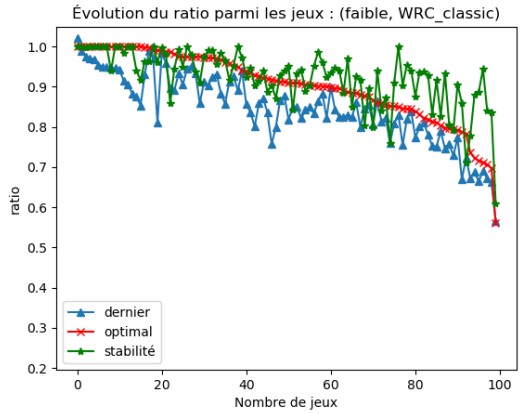
Tout d'abord, contrairement au cadre distribué où nous pouvons déduire que certains types de concession sont bien meilleurs que d'autres (*égalitaire*, *fort* et *Pareto*), cela n'est pas le cas dans le cadre décentralisé, où l'effet des types est plus discret. Nous pouvons néanmoins noter que deux types se démarquent légèrement, à savoir les types *égocentrique* et *utilitaire*. En effet, si nous prenons les résultats pour la stratégie WRC-Classic-Dec avec ces types, les pertes d'optimalité sont inférieures à 30% pour 70% des jeux, contre 60% des jeux pour les autres types. Avec la stratégie WRC-Coalitions-Dec, ces pertes d'optimalité inférieures à 30% sont concernent respectivement 70% et 65% des jeux pour les types *égocentrique* et *utilitaire*, tandis que pour les autres types, cela concerne uniquement 40



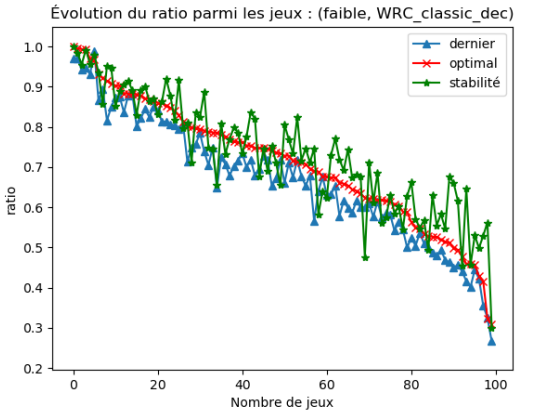
(a) Égocentrique / Dist.



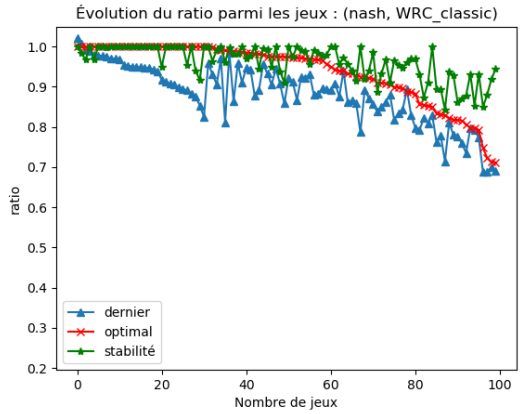
(b) Égocentrique / Dec.



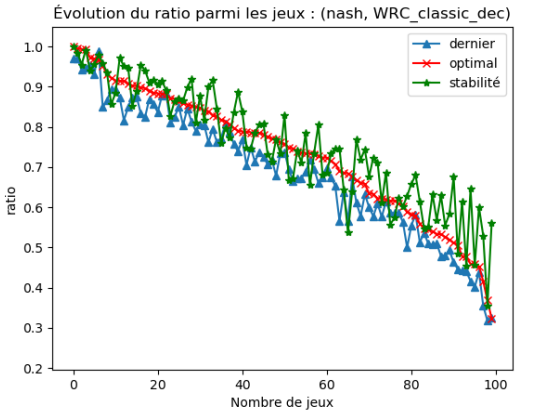
(c) Faible / Dist.



(d) Faible / Dec.

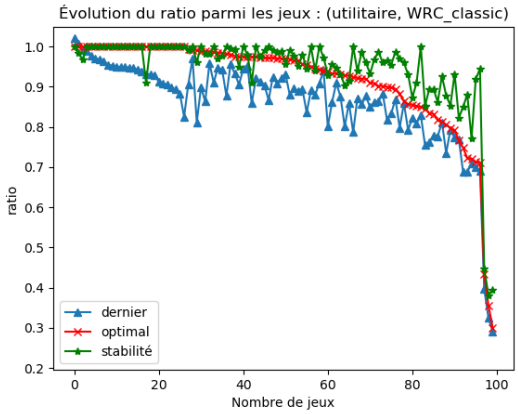


(e) Nash / Dist.

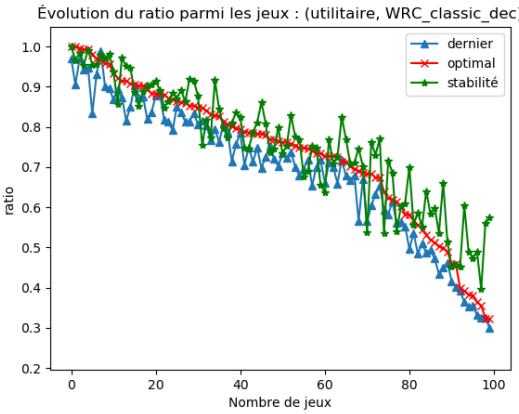


(f) Nash / Dec.

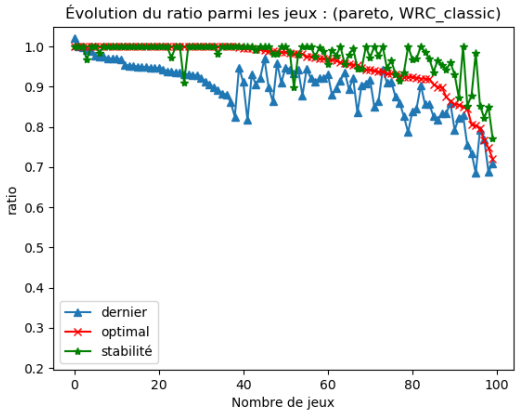
FIGURE 5.4 – Résultats pour les stratégies WRC-Classic(-Dec) selon le type de concession



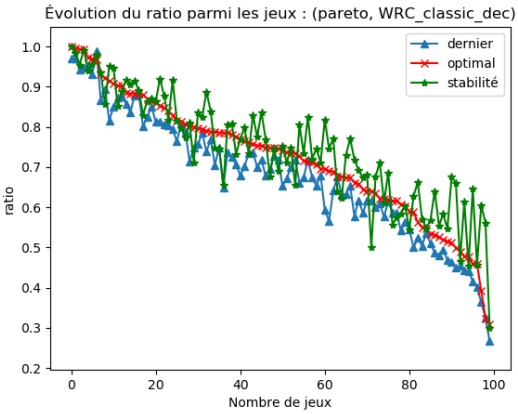
(a) Utillitaire / Dist.



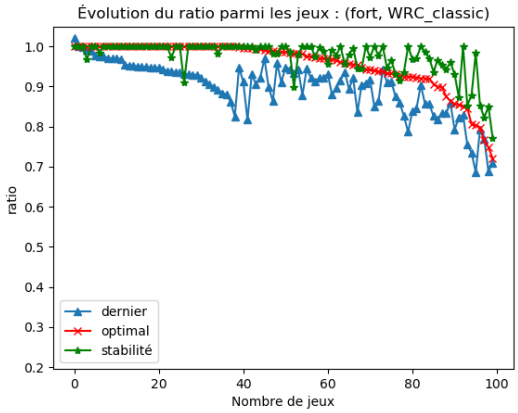
(b) Utillitaire / Dec.



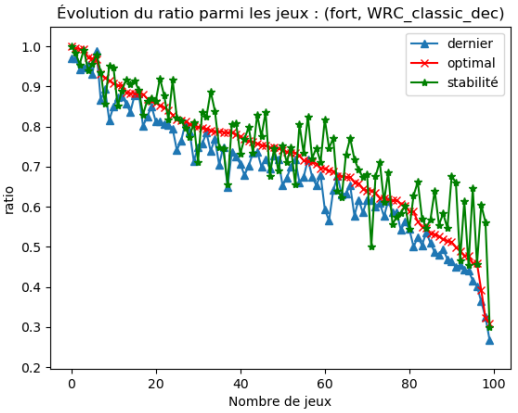
(c) Pareto / Dist.



(d) Pareto / Dec.

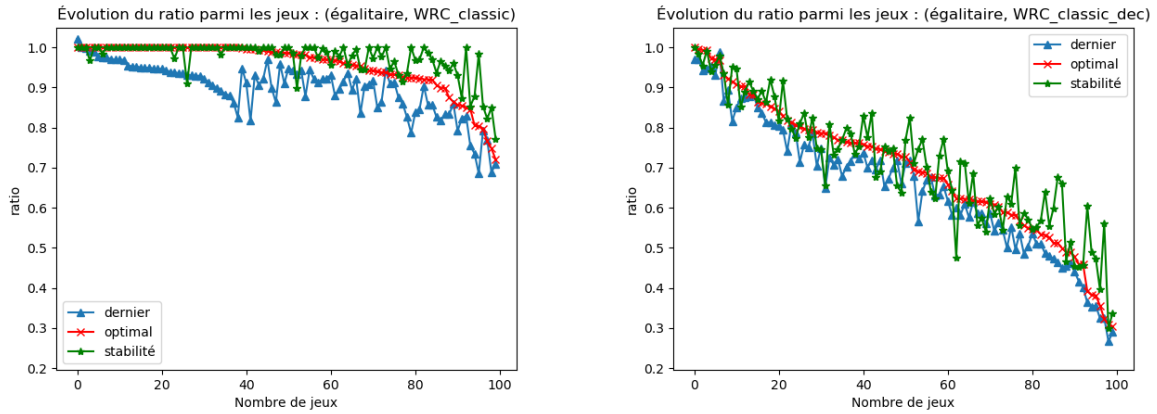


(e) Fort / Dist.



(f) Fort / Dec.

FIGURE 5.5 – Résultats pour les stratégies WRC-Classic(-Dec) selon le type de concession



(a) Égalitaire / Dist.

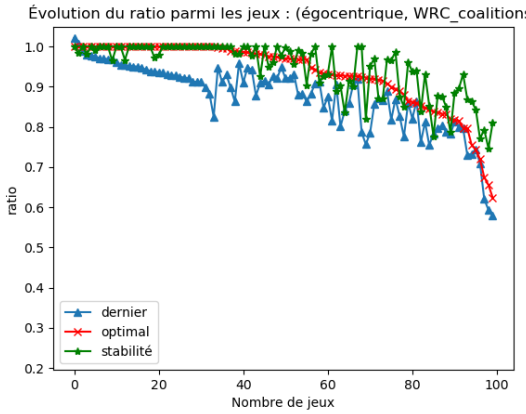
(b) Égalitaire / Dec.

FIGURE 5.6 – Résultats pour les stratégies WRC-Classic(-Dec) selon le type de concession

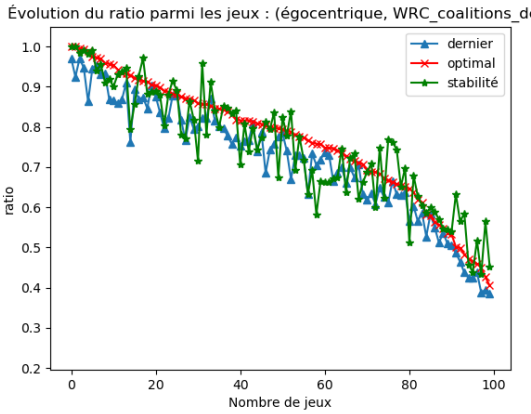
à 50% des jeux, le minimum étant avec le type *égalitaire*. Il est donc bien visible que les pertes d'optimalité sont bien plus importantes que dans le cadre distribué. Pour la stratégie WRC-Classic-Dec, seuls les types *égocentrique* et *utilitaire* atteignent un ratio de 1 (donc le maximum atteignable) pour 2 jeux, le reste des types ne l'atteignant que pour 1. Sur ce point, la stratégie WRC-Coalitions-Dec performe un peu plus avec la plupart des types, atteignant le ratio maximum avec 2 jeux sur tous les types hormis *utilitaire* et *égalitaire*. Pour les deux stratégies, les pertes maximales sont de l'ordre de 70 à 75%, sauf pour le type *égocentrique* conjointement à la stratégie WRC-Coalitions-Dec, où ces pertes maximales sont de 60%. Les résultats sont donc inverses au cadre distribué, où la stratégie WRC-Coalitions était meilleure que WRC-Classic, tandis que dans ce cadre décentralisé, WRC-Classic-Dec produit moins de pertes d'optimalité que WRC-Coalitions-Dec, sauf pour le type *égocentrique* qui donne des résultats proches pour les deux stratégies. Ce type est celui qui performe le mieux qu'importe la stratégie. Nous pouvons également conclure qu'une approche distribuée est plus appropriée qu'une approche décentralisée.

5.6 Conclusion

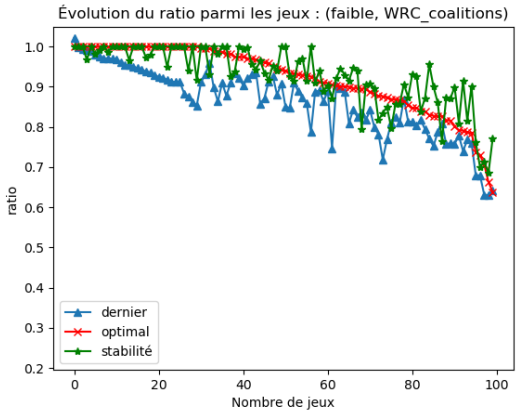
Nous avons proposé un protocole distribué et un décentralisé pour la formation de coalitions, fondés sur un protocole de négociations monotones, pour lesquels nous avons également proposé de nouvelles stratégies de concessions. Nous avons montré que ces stratégies adaptées à la notion de coalitions sont plus efficaces dans un contexte distribué.



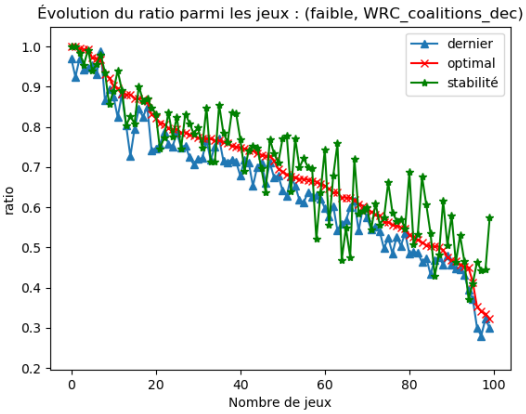
(a) Égocentrique / Dist.



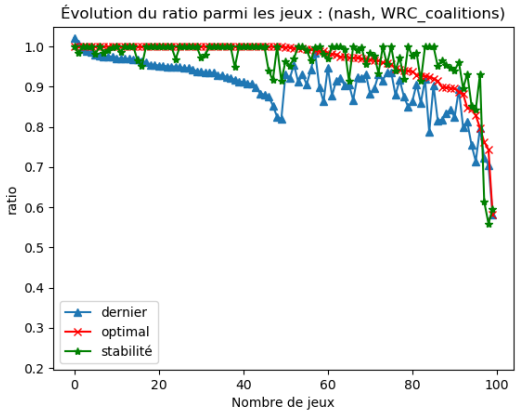
(b) Égocentrique / Dec.



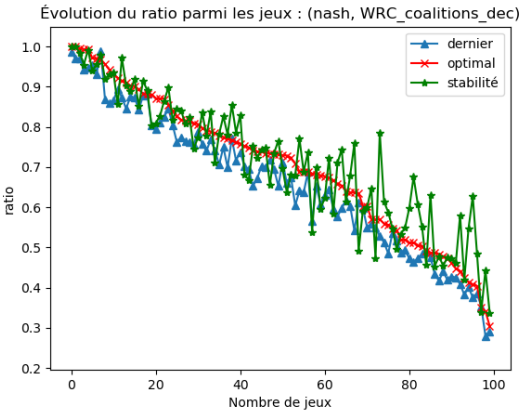
(c) Faible / Dist.



(d) Faible / Dec.

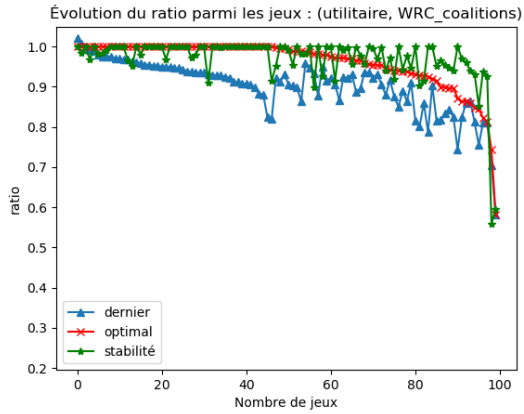


(e) Nash / Dist.

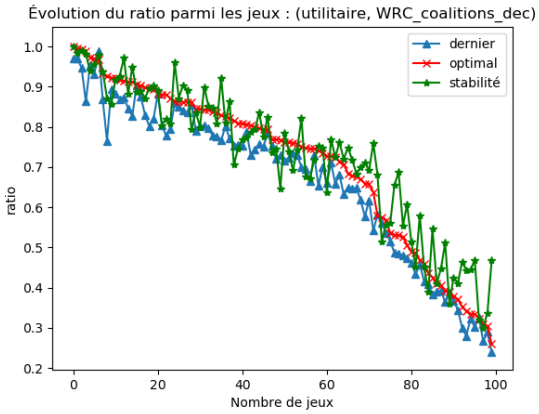


(f) Nash / Dec.

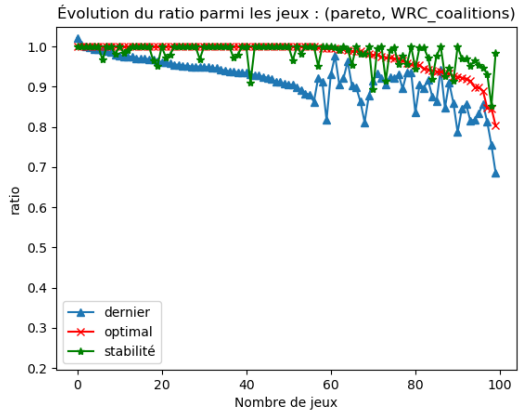
FIGURE 5.7 – Résultats pour WRC-Coalitions(-Dec) selon le type de concession



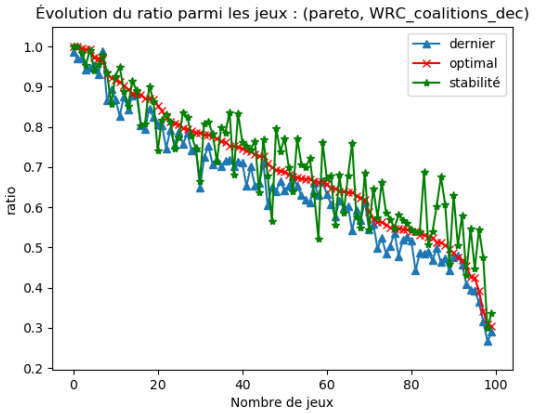
(a) Utillitaire / Dist.



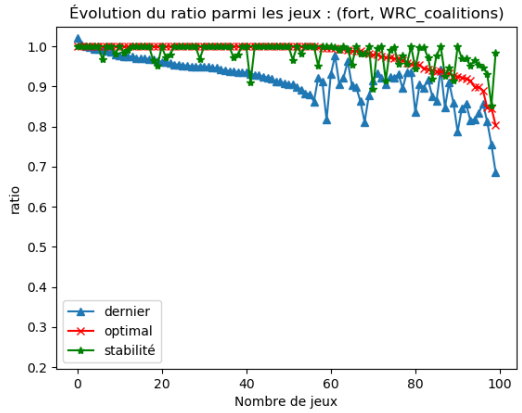
(b) Utillitaire / Dec.



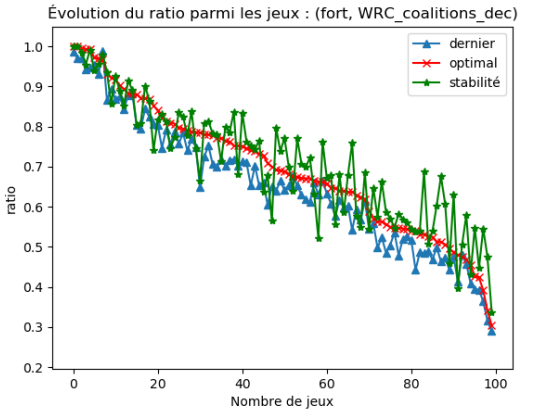
(c) Pareto / Dist.



(d) Pareto / Dec.

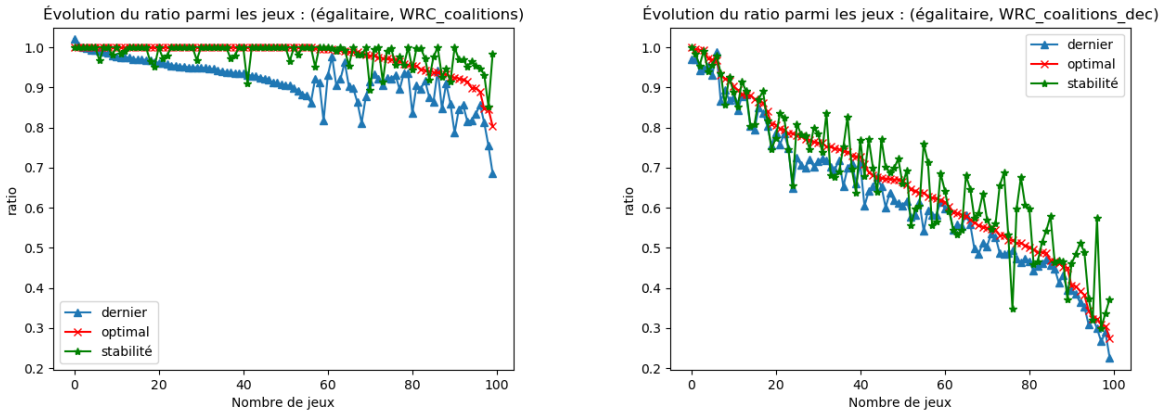


(e) Fort / Dist.



(f) Fort / Dec.

FIGURE 5.8 – Résultats pour WRC-Coalitions(-Dec) selon le type de concession



(a) Égalitaire / Dist.

(b) Égalitaire / Dec.

FIGURE 5.9 – Résultats pour WRC-Coalitions(-Dec) selon le type de concession

La comparaison entre les approches distribuée et décentralisée sont clairement en faveur de la première, qui est bien plus efficace, notamment avec le type de concession égalitaire. En effet, les pertes d'optimalité engendrées par la décentralisation sont très grandes, que ce soit en termes de valeur absolue, ou bien en nombre de jeux résolus efficacement. Tandis que dans le meilleur des cas, le protocole distribué trouve des solutions optimales dans 60% des jeux, la version décentralisée ne le permet que dans 1 à 2% des jeux, ce qui est bien trop faible. Dans le cadre décentralisé, les types de concessions n'entraînent pas de différence majeure, hormis le type *égocentrique* qui devient le plus performant, la meilleure stratégie tout type confondu étant WRC-Classic-Dec en raison des pertes moindres qu'elle engendre.

Chapitre 6

Protocole stochastique de formation de coalitions basé sur des concessions

Sommaire

Le chapitre 5 a été consacré à la levée de l'hypothèse de centralisation et de connaissance commune au sein de la formation de coalitions, avec l'adaptation d'un protocole de négociations multilatérales fondé sur des concessions monotones. Nous en avons conclu qu'une approche distribuée est plus adaptée qu'une approche décentralisée, cette dernière provoquant des pertes attendues d'optimalité. Nous avons cependant souligné que les meilleurs types et stratégies de concession ne sont pas les mêmes dans les deux cadres,

et notamment le fait qu'une approche égocentrique est plus efficace dans le cadre décentralisé. D'autres hypothèses que nous souhaitions lever sont celles de la connaissance *a priori* des utilités des coalitions par les agents ainsi que la nature déterministe de ces utilités. Dans cette optique, nous avons proposé un modèle de formation de coalitions stochastique répétée dans le chapitre 3, mais également deux concepts de solutions fondés sur l'exploration. Ces concepts ont été mis en concurrence contre une approche gloutonne, qui, malgré la stochasticité, est plus performante. Nous souhaitons donc aborder à nouveau ces hypothèses par le prisme du protocole de concessions monotones adapté à la formation de coalitions. La première problématique de ce chapitre consiste donc en l'utilisation du protocole de concessions monotones adapté à la formation de coalitions mais dans un cadre stochastique et répété de cette dernière. Enfin, malgré les pertes d'optimalité engendrées dans la version décentralisée du protocole par rapport à sa version distribuée, une deuxième problématique sera d'unir les diverses problématiques traitées dans ce mémoire dans un même modèle et protocole, en abordant à la fois les aspects décentralisé, stochastique et répété de la formation de coalitions.

6.1 Un protocole distribué et stochastique

Dans un premier temps, la problématique abordée sera celle de l'adaptation à un cadre stochastique et répété du protocole de concessions monotones pour la formation de coalitions. Cela consiste en l'adaptation de certaines définitions d'éléments du protocole afin que ce dernier soit applicable à un jeu de coalitions stochastique répété, modèle que nous avons défini en section 3.1. Cependant, contrairement aux travaux menés sur la résolution de ces jeux, le protocole n'inclut pas de concepts de solutions, mais uniquement des stratégies de concession. Il faut donc adapter ces dernières dans ce cadre stochastique et répété, ce qui passe par la définition de biais d'exploration.

6.1.1 Protocole distribué dans un cadre stochastique répété

Le protocole de concessions monotones adapté à la formation de coalitions que nous avons défini dans le chapitre précédent doit désormais être applicable aux jeux de coalitions stochastique répété.

6.1.1.1 Fonction caractéristique stochastique

Nous avons défini les jeux de coalitions stochastiques répétés dans le chapitre 3, avec la définition 3.1. Pour rappel, ces jeux possèdent une fonction caractéristique réelle inconnue des agents, et une fonction caractéristique estimée, qui est la croyance des agents sur la première. Étant donné que nous nous plaçons dans un cadre distribué, il n'est pas nécessaire que les agents possèdent des connaissances individuelles, et donc peuvent partager la connaissance commune de cette fonction caractéristique estimée. La méthode d'estimation utilisée est celle à partir d'une connaissance *a priori*, comme décrite par la définition 3.4.

Nous nous plaçons donc dans ce cadre où les utilités des fonctions caractéristiques sont définies par des variables aléatoire suivant des lois normales, que les agents doivent apprendre en observant les utilités produites à chaque répétition.

6.1.1.2 Ajout de la répétition du protocole

Dans la section 3.1.2, nous avons défini deux méthodes d'estimation que les agents peuvent utiliser afin de mettre à jour la fonction caractéristique estimée. Ces méthodes s'appuient respectivement sur une connaissance *a priori* et sur un apprentissage par inférence. La méthode d'estimation sur une connaissance *a priori* ayant fait ses preuves en termes de regret dans les expérimentations du chapitre 4, et possédant une meilleure connaissance *a priori* grâce à l'initialisation d'une fonction caractéristique estimée suivant la même classe que la réelle, nous faisons le choix ici d'utiliser cette méthode.

La répétition étant déjà un élément des jeux de coalitions stochastiques répétés, nous n'avons pas besoin de répéter le protocole à proprement parler, mais seulement de l'appliquer à chaque répétition du jeu, comme si c'était un nouveau protocole. Toutefois, la solution donnée par le protocole fournira les observations au jeu de coalitions stochastique répété, que les agents utiliseront pour mettre à jour leurs croyances.

6.1.2 Adaptation au contexte stochastique

La fonction caractéristique réelle étant inconnue des agents, qui en font une estimation, ces derniers doivent désormais formuler leurs propositions non plus sur une connaissance certaine mais sur les estimations effectuées. Il est cependant fait l'hypothèse ici que ces estimations sont communes à l'ensemble des agents, qui partagent donc leurs connaissances, tel que le cadre distribué l'exige. Toutefois, une problématique reste la même que

pour l'apprentissage par renforcement : pouvons-nous trouver un équilibre exploration-exploitation minimisant le regret des agents ? Pour aborder cette problématique, nous utiliserons à nouveau le biais d'exploration UCB pour les coalitions défini en section 3.2.1. Ensuite, nous pourrons proposer de nouvelles stratégies pour le protocole couplées à de l'exploration.

6.1.2.1 Propositions et règle de distribution

En raison de l'ajout de la stochasticité et de la répétition, la règle de distribution doit être légèrement adaptée. En effet, celle-ci doit être définie en fonction des croyances des agents à un pas de temps précis. Pour cela, il faut également adapter les définitions de surplus et de part de surplus à la fonction caractéristique estimée, tel que définies ci-dessous.

Définition 6.1 (Surplus *ex-ante*)**.** *Le* surplus *ex-ante* $\hat{S}_C$ *d'une coalition* C *est :*

$$\hat{S}_C = \hat{v}(C) - \sum_{a \in C} \hat{v}(a)$$

Définition 6.2 (Part de surplus *ex-ante*)**.** *La* part de surplus *ex-ante* $\hat{S}_C^{a_i}$ *d'un agent* a_i *dans sa coalition* C*, est le surplus de* C *divisé par son nombre d'agents :*

$$\hat{S}_C^{a_i} = \frac{\hat{S}_C}{|C|}$$

Ainsi, la règle de distribution, qui distribue toujours le surplus de façon égalitaire, est définie comme suit.

Définition 6.3 (Règle de distribution stochastique)**.** *Le gain de l'agent* a_i *appartenant à la coalition* C *avec un surplus* $\hat{S}_C > 0$ *est défini comme :*

$$\hat{x}_{a_i}^C = \hat{v}(\{a_i\}) + \hat{S}_C^{a_i}$$

Les propositions *ex-ante* sont donc construites à partir de l'estimation sur des connaissances *a priori*.

Définition 6.4 (Proposition *ex-ante*)**.** *Étant donné un jeu* $\mathcal{G}$*, une* proposition *de l'agent* a_i *au pas de temps* t*, notée* p_i^t*, est une solution* $S_{ante}^t = \langle \mathcal{CS}^t, \vec{x}_{ante}^t \rangle$ *où* $\vec{x}_{ante}^t$ *est un vecteur de gains* $\langle \hat{x}_{a_j}^{C_j^i} \rangle$ *où* C_j^i *est la coalition de l'agent* a_j *dans la proposition* p_i^t*.*

La notion d'*accord commun* pour des propositions *ex-ante* est la même que pour le cadre déterministe. La notion de conflit ne change pas. Une fois qu'une proposition *ex-ante* fait l'objet d'un accord commun, les coalitions proposées sont formées et le gain réel des agents est calculé grâce à la même règle de distribution, mais appliquée sur l'utilité réelle produite par la coalition. Si cette utilité ne permet pas d'obtenir un surplus positif, une part égalitaire du surplus négatif est retirée à chaque agent sur le montant de l'utilité de leur coalition singleton qui leur avait été réservé.

Exemple 36. *Prenons un exemple d'une coalition* $C = \{a_1, a_2\}$ *de taille* 2*. Si l'estimation de l'utilité de* C *est telle que* $\hat{v}(C) = 3.5$ *et celles des coalitions singletons telles que* $\hat{v}(\{a_1\}) = 2$; $\hat{v}(\{a_2\}) = 1$*. Le surplus* ex-ante *est* $\hat{S}_C = 0.5$*, le vecteur de gains* ex-ante *pour cette coalition est* $\vec{x}^t_{ante} = \{2.25, 1.25\}$*. Si la coalition produit un gain réel de* 2.5*, alors les agents percevront l'utilité de leur coalition singleton, plus une part de surplus réel (négatif)* $S_C = -0.5$*, tel que* $\vec{x}^t = \{1.75, 0.75\}$.

6.1.2.2 Biais d'exploration

Afin d'intégrer une notion d'exploration dans le protocole, il semble pertinent de le faire sur les stratégies de concession. En effet, ces stratégies permettent aux agents de négocier les propositions, et si une de ces dernières mérite d'être explorée, c'est au niveau des stratégies que les agents pourront le décider. Cependant, nous devons définir sous quelle forme cette exploration sera intégrée. Tout comme nous l'avons fait dans le chapitre 3, c'est sur l'analogie entre la formation de coalitions stochastique répétée et les bandits manchots que nous nous appuierons. Nous avons défini en section 3.2.1 un *biais d'exploration coalitionnel* (définition 3.8) fondé sur le terme d'exploration du biais UCB des bandits manchots, qui caractérise l'intérêt à former une coalition selon le nombre de fois où elle a été formée, par rapport au nombre total de coalitions formées. Nous réutilisons donc ce biais ici. Cependant, les expérimentations menées dans le chapitre 4 montrent que selon l'utilisation de ce biais, les résultats peuvent différer d'une façon non-négligeable, c'est pourquoi nous proposons quatre caractérisations pour l'utilisation de ces biais sur les propositions.

La première variante caractérise simplement l'intérêt exploratoire d'une proposition pour les agents d'une coalition donnée.

Définition 6.5 (Biais collectif)**.** *Soient une coalition* C_i *et une proposition* p_j*, le biais*

d'exploration collectif*, noté* UCB-C*, est défini comme suit :*

$$UCB\text{-}C(C_i, p_j, t) = \sum_{k \in C_i} \gamma_{coal}(C_k^{p_j}, t)$$

Tout comme il avait été argumenté en section 3.2.1, les agents sont souvent amenés à comparer des structures différentes, et donc des coalitions composées de membres différents, et donc potentiellement prendre en compte l'intérêt à explorer d'autres agents non impliqués. De plus, dans la définition précédente, il est simple de voir que si deux agents de la coalition C_i se retrouvent dans la même coalition dans la proposition p_j, alors le biais d'exploration coalitionnel pour cette coalition sera compté deux fois. C'est pourquoi, à l'instar de ce qui a été fait en section 3.2.1, nous définissons un biais caractérisant une notion d'exploration individuelle.

Définition 6.6 (Biais individuel)**.** *Soient une coalition C_i et une proposition p_j, le biais d'exploration* individuel*, noté* UCB-I*, est défini comme suit :*

$$UCB\text{-}I(C_i, p_j, t) = \sum_{k \in C_i} \frac{\gamma_{coal}(C_k^{p_j}, t)}{|C_k^{p_j}|}$$

La variante suivante caractérise l'exploration moyenne qu'une proposition produit, selon le nombre de coalitions que forment les agents d'une coalition donnée dans la proposition.

Définition 6.7 (Biais moyen)**.** *Soient une coalition C_i et une proposition p_j, le biais d'exploration* moyen*, noté* UCB-M*, est défini comme suit :*

$$UCB\text{-}M(C_i, p_j, t) = \frac{\sum_{k \in C_i} \gamma_{coal}(C_k^{p_j}, t)}{|CS_{C_i}(p_j)|}$$

où $CS_{C_i} = \{C_k, \forall k \ni C_i | p_j\}$

La dernière variante du biais d'exploration porte sur l'ensemble de la structure proposée dans une proposition. Cela est simplement la somme de l'ensemble des biais d'exploration coalitionnels pour cette structure, divisé par le nombre d'agents du jeu.

Définition 6.8 (Biais structurel)**.** *Soient une coalition et une proposition p_j, le biais d'exploration* structurel*, noté* UCB-S*, est défini comme suit :*

$$UCB\text{-}S(C_i, p_j, t) = \frac{\sum_{C_k \in p_j} \gamma_{coal}(C_k, t)}{|N|}$$

Remarque. *Bien que la coalition fasse partie des paramètres de ce dernier biais, elle n'a strictement aucun impact dans celui-ci. Nous la laissons uniquement dans un but de généralisation des notations dans les stratégies à venir.*

6.1.2.3 Stratégies de concession

Nous définissons dans la suite de nouvelles stratégies pour ce cadre stochastique, presque toutes étant fondées sur la stratégie la plus efficace du cadre distribué, *WRC-Coalitions*, et une dernière fondée sur *WRC-Classic* (voir section 5.1.3.2). Les noms des nouvelles stratégies ne spécifient donc pas le mot "Coalitions", tandis que nous le spécifions pour la stratégie fondée sur *WRC-Classic*. De plus, notre définition du biais d'exploration structurel nous permet de proposer des stratégies de concession spécifiques à ce biais. Remarquons que nous proposons d'intégrer le biais d'exploration aux stratégies et non pas aux types de concessions car c'est avec la stratégie de concession que les agents peuvent se mettre d'accord sur le fait ou non d'explorer, alors que sur le type de concession seul l'agent faisant la concession ne le prendra en compte. Lorsqu'une stratégie peut utiliser n'importe quel biais, nous utilisons la notation UCB-*.

Une première stratégie consiste à intégrer à la stratégie WRC-Coalitions un terme d'exploration similaire au calcul de la valeur Z enrichi d'un biais UCB. Cela forme donc un terme décrivant la perte minimale d'exploration d'un agent s'il concède. Afin de garder la sémantique initiale de la stratégie WRC, avec laquelle la valeur 1 est considérée comme maximale et donc qu'un agent ne concède pas, nous divisons par deux la somme du terme WRC-Coalitions et du nouveau terme.

Définition 6.9 (WRC-Additional-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{\left[\frac{x_{C_i}(p_i) - \min\limits_{\forall j \in N} x_{C_i}(p_j)}{x_{C_i}(p_i)}\right] + \left[\frac{UCB\text{-}*(C_i,p_i,t) - \min\limits_{\forall k \in N} UCB\text{-}*(C_i,p_k,t)}{UCB\text{-}*(C_i,p_i,t)}\right]}{2}$$

Pour la stratégie suivante, nous intégrons également le nouveau terme décrit précédemment à WRC-Coalitions, mais en tant que multiplicateur, afin que ce terme d'exploration pondère la valeur Z de WRC-Coalitions. Ainsi, la valeur Z maximale reste 1.

Définition 6.10 (WRC-Proportionalized-UCB)**.** *L'agent qui concède est l'agent a_i pour*

qui la valeur Z_{a_i} est la plus petite, où :

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall j \in N} x_{C_i}(p_j)}{x_{C_i}(p_i)} \times \frac{UCB\text{-}*(C_i, p_i, t) - \min\limits_{\forall k \in N} UCB\text{-}*(C_i, p_k, t)}{UCB\text{-}*(C_i, p_i, t)}$$

Là où les stratégies précédentes séparent explicitement le valeur de la coalition et son biais d'exploration en tant que terme additionnel ou multiplicateur de la valeur Z, la stratégie suivante est une stratégie WRC-Coalitions où les gains et le biais sont directement additionnés avant de calculer la valeur Z.

Définition 6.11 (WRC-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{[x_{C_i}(p_i) + UCB\text{-}*(C_i, p_i, t)] - \min\limits_{\forall j \in N} [x_{C_i}(p_j) + UCB\text{-}*(C_i, p_j, t)]}{x_{C_i}(p_i) + UCB\text{-}*(C_i, p_i, t)}$$

Les stratégies suivantes n'utilisent que le biais d'exploration structurel *UCB-S*. Ainsi, l'exploration considérée ne provient que de la proposition courante de l'agent pour lequel nous calculons la valeur Z. La première stratégie consiste juste en l'addition du biais d'exploration structurel au nominateur et au dénominateur du calcul de WRC-Coalitions.

Définition 6.12 (WRC-Fractional-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall j \in N} x_{C_i}(p_j) + UCB\text{-}S(C_i, p_i, t)}{x_{C_i}(p_i) + UCB\text{-}S(C_i, p_i, t)}$$

La deuxième stratégie ne contenant que *UCB-S* est directement inspirée de la stratégie UCB des bandits manchots, où un terme d'exploration UCB est additionné à un terme d'exploitation. Ce dernier est donc ici le calcul de WRC-Coalitions, auquel nous ajoutons le biais *UCB-S*. Ainsi, chaque agent considérera l'ensemble des coalitions proposées dans son exploration.

Définition 6.13 (WRC-Bandit-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall j \in N} x_{C_i}(p_j)}{x_{C_i}(p_i)} + UCB\text{-}S(C_i, p_i, t)$$

Une dernière stratégie adaptée de WRC-Coalitions est similaire à WRC-Additional-UCB sans la division, et où le type de biais UCB est fixé. Cela est fait de sorte à ce que le maximum du terme soit désormais de 2, avec un terme d'exploration le plus général possible, c'est-à-dire ne dépendant pas d'une coalition précise.

Définition 6.14 (WRC-Superadditive-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \left[\frac{x_{C_i}(p_i) - \min\limits_{\forall j \in N} x_{C_i}(p_j)}{x_{C_i}(p_i)}\right] + \left[\frac{\textit{UCB-S}(C_i, p_i, t) - \min\limits_{\forall k \in N} \textit{UCB-S}(C_i, p_k, t)}{\textit{UCB-S}(C_i, p_i, t)}\right]$$

Enfin, une dernière stratégie est basée sur la stratégie distribuée WRC-Classic et non WRC-Coalitions. Elle est l'équivalent de WRC-UCB mais fondée sur WRC-Classic et non sur WRC-Coalitions. Elle n'utilise également que le biais d'exploration structurel *UCB-S* car celui-ci est le seul biais ne dépendant pas d'une coalition précise mais uniquement d'une structure.

Définition 6.15 (WRC-**Classic**-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{[x_i(p_i) + \textit{UCB-S}(C_i, p_i, t)] - \left[\min\limits_{\forall j \in N} x_i(p_j) + \textit{UCB-S}(C_i, p_j, t)\right]}{x_i(p_i) + \textit{UCB-S}(C_i, p_i, t)}$$

6.1.2.4 Étapes du protocole

Les éléments du protocole étant désormais définis et adaptés, nous pouvons détailler les tours de négociation. Ici, l'ensemble des agents connaissent la fonction caractéristique estimée, et chaque agent possède une liste noire privée servant à mémoriser les coalitions et structures de coalitions ayant été proposées et rejetées, et celles qui ont été rejetées car ne satisfaisant pas le type de concession.

1. Chaque agent calcule le surplus *ex-ante* des coalitions dont il peut faire partie,
2. Chaque agent crée une liste noire privée de coalitions et structures de coalitions, destinée à contenir les structures rejetées que ce soit par concession ou car ne respectant pas le type de concession. Les coalitions appartenant à des structures ayant toutes été ajoutées à la liste noire sont elles-mêmes ajoutées dans cette dernière.

3. Au premier tour, chaque agent fait une proposition *ex-ante* initiale en choisissant sa coalition, notée C^*, parmi celles qui maximisent sa part de surplus *ex-ante*, puis en choisissant la structure de coalitions qui maximise le bien-être social estimé, notée $\mathcal{CS}^*$, et qui inclut la coalition choisie C^*,
4. À chaque tour suivant, chaque agent garde sa proposition ou concède, selon sa stratégie de concession,
5. Si un agent concède, la structure de coalitions précédente $\mathcal{CS}^*$ est ajoutée à la liste noire, et l'agent essaye de construire une nouvelle proposition *ex-ante* satisfaisant son type de concession avec une autre structure de coalitions $\mathcal{CS}^{*'}$ qui inclut également sa coalition choisie C^*. Si une structure $\mathcal{CS}^{*'}$ ne satisfait pas un type de concession, elle est ajoutée à la liste noire. S'il n'y a plus de structure de coalitions possible avec la coalition C^*, cette dernière est ajoutée à la liste noire, et l'agent choisit une autre coalition, $C^{*'}$, qui maximise sa part de surplus *ex-ante*. Un agent qui propose une coalition $C^{*'}$ dont le surplus *ex-ante* est égal à 0 se retire du processus car il ne pourra pas gagner quoique ce soit.
6. Répéter à partir de l'étape (4) jusqu'à ce qu'un accord soit atteint ou qu'aucun agent ne reste dans le processus, c'est-à-dire qu'aucun agent ne puisse faire une proposition *ex-ante* où il gagnerait quelque chose (donc un *conflit*).
7. Formation des coalitions formées, calcul de la répartition des gains selon la même règle de distribution et ajout des observations à la liste d'observations du RSCG pour mise à jour des croyances. Les listes noires sont vidées.

Les différences avec le protocole distribué déterministe défini dans le chapitre 5 résident donc dans l'utilisation des propositions et du surplus *ex-ante*, ainsi que dans l'étape d'observation, de recalcul des gains, et de la mise à jour des croyances.

6.2 Un protocole décentralisé et stochastique

La levée des hypothèses de connaissances *a priori* et de déterminisme de ces utilités ayant été traitée, nous souhaitons désormais unifier les différents protocoles traitant des hypothèses levées. Cela se fait par la fusion du protocole stochastique précédemment défini dans ce chapitre avec le protocole décentralisé défini au chapitre 5.

6.2.1 Adaptation des concepts décentralisés à la stochasticité

Tout comme pour la décentralisation de la version déterministe du protocole, la décentralisation de la version stochastique repose sur l'abandon de la connaissance commune des agents, nous devons à nouveau redéfinir la fonction caractéristique de sorte à ce que les agents ne connaissent que les utilités qui les concernent, et donc les coalitions dont ils font partie. Cette redéfinition de la fonction caractéristique implique donc la redéfinition des propositions.

6.2.1.1 Fonction caractéristique

Afin de réduire les connaissances des agents, nous redéfinissons simplement la fonction caractéristique par une restriction de cette dernière. Les agents n'auront donc qu'une connaissance partielle et individuelle de la fonction caractéristique stochastique (voir définition 3.1), ne portant que sur les coalitions dont il fait partie.

Définition 6.16 (Fonction caractéristique stochastique individuelle)**.** *Soit v_i la* fonction caractéristique stochastique individuelle *de l'agent a_i. Celle-ci est définie comme étant la restriction de la fonction caractéristique v d'un jeu $\mathcal{G}$ à G_i l'ensemble de coalitions auxquelles a_i appartient, i.e. :*

$$G_i = \{C \mid \forall C \in 2^N \ t.q.\ a_i \in C\}$$

v_i est donc définie comme suit :

$$v_i : G_i \to \mathcal{X}^{G_i}$$

Il est important de noter que la fonction caractéristique v est inconnue des agents, c'est pourquoi chaque agent possède une estimation $\hat{v}$ de sa fonction caractéristique stochastique individuelle, qu'il met à jour à partir d'une estimation sur une connaissance *a priori* (voir définition 3.4).

Exemple 37. *Soit la fonction caractéristique v définie dans l'exemple 31 (page 116). La fonction caractéristique stochastique individuelle de*  *est :*

$$v_{\text{  }} = \Big\{\{\text{  }\} = \mathcal{N}(0.4, 0.1)\ ;\ \{\text{  }\} = \mathcal{N}(0.6, 0.1)\ ;$$
$$\{\text{  }\} = \mathcal{N}(0.7, 0.2)\ ;\ \{\text{  }\} = \mathcal{N}(0.7, 0.3)\Big\}$$

Cette fonction $v_{\text{  }}$ est inconnue de l'agent.

6.2.1.2 Propositions

Les propositions *ex-ante* doivent être redéfinies afin que les agents ne proposent que la coalition qu'ils souhaitent former. Le vecteur de gain proposé est construit à partir de la fonction caractéristique stochastique indiviudelle estimée de l'agent faisant la proposition, avec la règle de distribution donnée par la définition 6.3.

Définition 6.17 (Proposition *ex-ante* décentralisée)**.** *Étant donné un jeu $\mathcal{G}$, une* proposition ex-ante *dans un cadre* décentralisé *de l'agent a_i, notée p_i, est un tuple $S^t_{dec} = \langle C^t, \vec{x}^t_{ante} \rangle$ où $\vec{x}^t_{ante}$ est un vecteur de gains $\langle \hat{x}^{C^i_j}_{a_j} \rangle$ où C^i_j est la coalition de l'agent a_j dans la proposition p_i.*

La notion d'*accord local* donnée par la définition 5.17 est également valable pour les propositions *ex-ante* décentralisées. La notion de *conflit* reste identique. Une fois qu'une proposition *ex-ante* fait l'objet d'un accord local, la coalition proposée est formée et le gain réel des agents est calculé grâce à la même règle de distribution, mais appliquée sur l'utilité réelle produite par la coalition. Si cette utilité ne permet pas d'obtenir un surplus positif, une part égalitaire du surplus négatif est retirée à chaque agent sur le montant de l'utilité de leur coalition singleton qui leur avait été réservé (voir exemple 36).

6.2.2 Adaptation au contexte décentralisé

Outre les définitions actualisées précédemment, des stratégies de concession adaptées doivent être proposées. Pour cela, nous nous appuierons à nouveau sur le biais d'exploration UCB pour les coalitions défini en section 3.2.1.

6.2.2.1 Biais d'exploration

Les biais d'exploration définis en section 6.1.2.2 le sont de sorte à prendre en compte les propositions contenant des structures de coalitions. Toutefois, les stratégies que nous souhaitons définir désormais doivent porter sur des coalitions et non des structures. Nous devons donc adapter ces biais d'exploration afin de pouvoir les utiliser dans ce contexte décentralisé. À l'instar de ce qui a été fait pour la décentralisation du protocole décentralisé, lorsqu'un agent n'apparaît pas dans une proposition, nous le considérerons dans sa coalition singleton.

Définition 6.18 (Biais collectif)**.** *Soient une coalition C_i et une proposition p_j, le biais d'exploration* collectif *dans un cadre décentralisé, noté* UCB-C-Dec*, est défini comme*

suit :

$$UCB\text{-}C\text{-}Dec(C_i, p_j, t) = \sum_{a_k \in C_i \wedge a_k \in p_j} \gamma_{coal}(C_k^{p_j}, t) + \sum_{a_k \in C_i \wedge a_k \notin p_j} \gamma_{coal}(\{a_k\}, t)$$

Définition 6.19 (Biais individuel)**.** *Soient une coalition C_i et une proposition p_j, le biais d'exploration* individuel *dans un cadre décentralisé, noté* UCB-I-Dec*, est défini comme suit :*

$$UCB\text{-}I\text{-}Dec(C_i, p_j, t) = \sum_{a_k \in C_i \wedge a_k \in p_j} \frac{\gamma_{coal}(C_k^{p_j}, t)}{|C_k^{p_j}|} + \sum_{a_k \in C_i \wedge a_k \notin p_j} \gamma_{coal}(\{a_k\}, t)$$

Définition 6.20 (Biais moyen)**.** *Soient une coalition C_i et une proposition p_j, le biais d'exploration* moyen *dans un cadre décentralisé, noté* UCB-M-Dec*, est défini comme suit :*

$$UCB\text{-}M\text{-}Dec(C_i, p_j, t) = \frac{\sum_{a_k \in C_i \wedge a_k \in p_j} \gamma_{coal}(C_k^{p_j}, t) + \sum_{k \in C_i \wedge a_k \notin p_j} \gamma_{coal}(\{a_k\}, t)}{|CS_{C_i}(p_j)|}$$

où $CS_{C_i} = \{C_k\} \cup \{\{a_k\}, \forall a_k \in C_i \wedge a_k \notin p_j\}$

Définition 6.21 (Biais structurel)**.** *Soient une coalition et une proposition p_j, le biais d'exploration* structurel *dans un cadre décentralisé, noté* UCB-S-Dec*, est défini comme suit :*

$$UCB\text{-}S\text{-}Dec(C_i, p_j, t) = \frac{\gamma_{coal}(C_k, t) + \sum_{a_k \notin p_j} \gamma_{coal}(\{a_k\}, t)}{|N|}$$

6.2.2.2 Stratégies de concession

À l'instar de ce qui a été proposé dans le chapitre 5, nous fondons la différence entre les versions stochastiques distribuées et les versions stochastiques décentralisées des stratégies sur l'ensemble des agents pour lesquels les propositions sont considérées par la stratégie, à savoir l'ensemble **E** dans les définitions suivantes. Pour rappel, cet ensemble **E** décrit quelles propositions seront considérées par la stratégie afin de calculer la valeur de concession, en raison des connaissances restraintes des agents, ces derniers ne pouvant donc pas raisonner sur l'ensemble complet des propositions, mais uniquement celles d'intérêt pour eux. Formellement, dans les versions distribuées, cet ensemble **E** contient les propositions de tous les agents. Dans les versions décentralisées, cet ensemble $\mathbf{E}_i$ pour un agent a_i est composé des propositions p_j dans lesquelles il apparaît, telles que ces propositions sont différentes de celle de a_i. Si aucune proposition ne correspond à ces critères, alors l'agent

a_i intégrera dans son ensemble $\mathbf{E}_i$ la proposition suivante qu'il ferait en cas de concession, notée $p_i^{\rightarrow}$. La définition formelle est donc donnée par la définition 5.18.

Les stratégies proposées pour le cadre stochastique sont adaptées à ce cadre décentralisé, à savoir *WRC-Additional-UCB*, *WRC-Proportionalized-UCB*, *WRC-UCB*, *WRC-Fractional-UCB*, *WRC-Bandit-UCB*, *WRC-Superadditive-UCB*, et ce avec la même utilisation des différents types de biais (c'est-à-dire que certaines stratégies n'utilisent que le biais structurel). La stratégie *WRC-Classic-UCB* est également adaptée afin de conserver dans les expérimentations la sémantique individuelle de la stratégie *WRC-Classic*. Comme précédemment, *UCB-*-Dec* est la notation utilisée pour signifier que la stratégie peut utiliser n'importe quel biais d'exploration. Les trois premières stratégies sont dans ce cas.

Définition 6.22 (WRC-Additional-UCB-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{\left[\frac{x_{C_i}(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_{C_i}(p_j)}{x_{C_i}(p_i)}\right] + \left[\frac{\textit{UCB-*-Dec}(C_i,p_i,t) - \min\limits_{\forall p_k \in \boldsymbol{E}_i} \textit{UCB-*-Dec}(C_i,p_k,t)}{\textit{UCB-*-Dec}(C_i,p_i,t)}\right]}{2}$$

Définition 6.23 (WRC-Proportionalized-UCB-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_{C_i}(p_j)}{x_{C_i}(p_i)} \times \frac{\textit{UCB-*-Dec}(C_i,p_i,t) - \min\limits_{\forall p_k \in \boldsymbol{E}_i} \textit{UCB-*-Dec}(C_i,p_k,t)}{\textit{UCB-*-Dec}(C_i,p_i,t)}$$

Définition 6.24 (WRC-UCB-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{[x_{C_i}(p_i) + \textit{UCB-*-Dec}(C_i,p_i,t)] - \min\limits_{\forall p_j \in \boldsymbol{E}_i} [x_{C_i}(p_j) + \textit{UCB-*-Dec}(C_i,p_j,t)]}{x_{C_i}(p_i) + \textit{UCB-*-Dec}(C_i,p_i,t)}$$

Les stratégies suivantes n'utilisent que le biais structurel *UCB-S-Dec*, comme leurs versions distribuées.

Définition 6.25 (WRC-Fractional-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la*

valeur Z_{a_i} est la plus petite, où :

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_{C_i}(p_j) + \textit{UCB-S-Dec}(C_i, p_i, t)}{x_{C_i}(p_i) + \textit{UCB-S-Dec}(C_i, p_i, t)}$$

Définition 6.26 (WRC-Bandit-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{x_{C_i}(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_{C_i}(p_j)}{x_{C_i}(p_i)} + \textit{UCB-S-Dec}(C_i, p_i, t)$$

Définition 6.27 (WRC-Superadditive-UCB)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \left[\frac{x_{C_i}(p_i) - \min\limits_{\forall p_j \in \boldsymbol{E}_i} x_{C_i}(p_j)}{x_{C_i}(p_i)}\right] + \left[\frac{\textit{UCB-S-Dec}(C_i, p_i, t) - \min\limits_{\forall p_k \in \boldsymbol{E}_i} \textit{UCB-S-Dec}(C_i, p_k, t)}{\textit{UCB-S-Dec}(C_i, p_i, t)}\right]$$

Cette dernière stratégie est l'équivalent de *WRC-UCB-Dec* mais fondée sur *WRC-Classic-Dec* et non sur *WRC-Coalitions-Dec*.

Définition 6.28 (WRC-**Classic**-UCB-Dec)**.** *L'agent qui concède est l'agent a_i pour qui la valeur Z_{a_i} est la plus petite, où :*

$$Z_{a_i} = \frac{[x_i(p_i) + \textit{UCB-S-Dec}(C_i, p_i, t)] - \left[\min\limits_{\forall p_j \in \boldsymbol{E}_i} x_i(p_j) + \textit{UCB-S-Dec}(C_i, p_j, t)\right]}{x_i(p_i) + \textit{UCB-S-Dec}(C_i, p_i, t)}$$

6.2.2.3 Étapes du protocole

Le protocole pour cette version est donc un mélange du protocole décentralisé défini dans le chapitre 5, et le protocole stochastique défini dans la section précédente. Ici, les agents ne connaissent que leur propre fonction caractéristique stochastique individuelle, ainsi qu'une liste noire privée servant à mémoriser les coalitions ayant été proposées et rejetées, ou celles rejetées car ne satisfaisant pas le type de concession. De plus, il est fait l'hypothèse qu'au début du protocole, les agents s'échangent la connaissance de l'utilité de leurs coalitions singleton afin de pouvoir construire leurs propositions.

1. Chaque agent calcule le surplus *ex-ante* des coalitions dont il peut faire partie,

2. Chaque agent crée une liste noire privée de coalitions,
3. Au premier tour, chaque agent fait une proposition *ex-ante* initiale en choisissant sa coalition, notée C^*, parmi celles qui maximisent sa part de surplus *ex-ante*,
4. À chaque tour suivant, chaque agent garde sa proposition ou concède, selon sa stratégie de concession, sauf si un *accord local* existe, alors les agents concernés se retirent du processus, les autres agents répète le processus à partir de l'étape (3),
5. Si un agent concède, la coalition C^* est ajoutée à la liste noire, et l'agent choisit une autre coalition, $C^{*'}$, qui maximise sa part de surplus *ex-ante*. Un agent qui propose une coalition $C^{*'}$ dont le surplus *ex-ante* est égal à 0 se retire du processus car il ne pourra pas gagner quoique ce soit.
6. Répéter à partir de l'étape (4) jusqu'à ce que tous les agents possèdent un accord local (ce qui correspond donc à un accord total) ou qu'aucun agent ne reste dans le processus, c'est-à-dire qu'aucun agent ne puisse faire une proposition *ex-ante* où il gagnerait quelque chose (donc un *conflit*).
7. Formation des coalitions formées, calcul de la répartition des gains selon la même règle de distribution et ajout des observations à la liste d'observations du RSCG pour mise à jour des croyances. Les listes noires sont vidées.

La différence entre ce protocole décentralisé et le protocole décentralisé du chapitre 5 réside donc dans le fait que les propositions utilisées sont *ex-ante*, tout comme le surplus. Une autre différence est l'observation des utilités réelles, le recalcul des gains et la mise à jour des croyances.

6.3 Expérimentations et résultats : distribué stochastique contre distribué déterministe

Nous pouvons désormais mettre en application le protocole adapté à un cadre stochastique sur des jeux de coalitions stochastiques répétés. Nous pourrons observer l'effet du biais d'exploration sur les différentes stratégies, ainsi que les meilleurs couples de types et stratégies de concessions. Enfin, afin de juger l'utilité de l'ajout d'un biais d'exploration, nous utiliserons les stratégies définies dans le cadre distribué, et donc sans exploration, sur ces jeux, afin d'avoir une comparaison avec des stratégies gloutonnes.

6.3.1 Paramètres des expérimentations

Le tableau 6.1 récapitule les différentes configurations entre les stratégies avec les divers biais d'exploration. À noter la présence de deux stratégies du protocole définies dans le cadre distribué : *WRC-Coalitions* et *WRC-Classic* (voir section 5.1.3.2). L'ajout de ces stratégies nous permet d'avoir une comparaison à des stratégies n'intégrant pas d'exploration.

Stratégie	UCB-C	UCB-I	UCB-M	UCB-S	NO-UCB
WRC-Add.	✓	✓	✓	✓	✗
WRC-Prop.	✓	✓	✓	✓	✗
WRC-UCB.	✓	✓	✓	✓	✗
WRC-Frac.	✗	✗	✗	✓	✗
WRC-Bandit.	✗	✗	✗	✓	✗
WRC-Sup.	✗	✗	✗	✓	✗
WRC-Cl-UCB.	✗	✗	✗	✓	✗
WRC-Class.	✗	✗	✗	✗	✓
WRC-Coal.	✗	✗	✗	✗	✓

TABLE 6.1 – Récapitulatif des stratégies et biais paramétrant le protocole

6.3.2 Tableaux de résultats

Nous présentons ci-après les résultats des expérimentations sous forme de tableaux, qui montrent les moyennes et variance du ratio à l'optimal-protocole sur l'ensemble des pas de temps des 100 jeux pour chaque paramétrage possible du protocole, à savoir le type de concession pour les lignes, la stratégie de concession pour les colonnes, et si tel est le cas, le type de biais d'exploration UCB utilisé (montré par les sous-colonnes des stratégies). Nous choisissons de montrer uniquement le ratio à l'optimal-protocole car elle est la plus représentative, un ratio de 1 étant le meilleur résultat possible. Ainsi, le tableau 6.2 montre les résultats pour tout type de concessions et les stratégies WRC-Additional-UCB et WRC-Classic-UCB, le tableau 6.3 pour les stratégies WRC-Proportionalized-UCB et WRC-Fractional-UCB, le tableau 6.4 pour les stratégies WRC-UCB et WRC-Bandit-UCB, et enfin le tableau 6.5 pour les stratégies WRC-Superadditive-UCB, WRC-Classic et WRC-Coalitions.

La figure 6.2 en page 167 donne le tableau de résultats complet. La ligne étiquettée *Total* de cette figure représente le nombre de type de concession dont les meilleurs résultats

	Additionnal				Classic_UCB
	C	I	M	S	S
Egalitaire	(0.88, 0.01)	(0.86, 0.02)	(0.87, 0.01)	(0.88, 0.01)	(0.86, 0.01)
Égocentrique	(0.77, 0.02)	(0.72, 0.03)	(0.76, 0.02)	(0.75, 0.03)	(0.74, 0.02)
Nash	(0.83, 0.01)	(0.81, 0.02)	(0.84, 0.01)	(0.83, 0.02)	(0.80, 0.02)
Pareto	(0.79, 0.02)	(0.79, 0.02)	(0.81, 0.02)	(0.79, 0.02)	(0.77, 0.02)
Fort	(0.45, 0.04)	(0.45, 0.05)	(0.45, 0.04)	(0.45, 0.04)	(0.43, 0.04)
Utilitaire	(0.86, 0.01)	(0.84, 0.02)	(0.86, 0.01)	(0.85, 0.01)	(0.83, 0.01)
Faible	(0.58, 0.03)	(0.57, 0.04)	(0.60, 0.03)	(0.59, 0.03)	(0.54, 0.03)

TABLE 6.2 – Résultats pour WRC-Additional-UCB et WRC-Classic-UCB

	Propotionalized				Fractional
	C	I	M	S	S
Egalitaire	(0.75, 0.03)	(0.64, 0.03)	(0.86, 0.02)	(0.66, 0.03)	(0.82, 0.02)
Égocentrique	(0.72, 0.02)	(0.68, 0.03)	(0.76, 0.03)	(0.69, 0.03)	(0.71, 0.02)
Nash	(0.72, 0.03)	(0.64, 0.04)	(0.83, 0.02)	(0.66, 0.04)	(0.78, 0.02)
Pareto	(0.65, 0.04)	(0.55, 0.04)	(0.81, 0.02)	(0.56, 0.04)	(0.74, 0.02)
Fort	(0.37, 0.04)	(0.30, 0.02)	(0.45, 0.05)	(0.34, 0.03)	(0.45, 0.04)
Utilitaire	(0.74, 0.03)	(0.64, 0.04)	(0.83, 0.02)	(0.68, 0.04)	(0.81, 0.02)
Faible	(0.52, 0.04)	(0.42, 0.04)	(0.61, 0.03)	(0.45, 0.04)	(0.53, 0.03)

TABLE 6.3 – Résultats pour WRC-Proportionalized-UCB et WRC-Fractional-UCB

	UCB				Bandit
	C	I	M	S	S
Egalitaire	(0.86, 0.01)	(0.80, 0.02)	(0.86, 0.01)	(0.87, 0.01)	(0.87, 0.01)
Égocentrique	(0.76, 0.02)	(0.73, 0.02)	(0.77, 0.02)	(0.75, 0.03)	(0.76, 0.03)
Nash	(0.83, 0.01)	(0.79, 0.02)	(0.84, 0.01)	(0.81, 0.02)	(0.83, 0.02)
Pareto	(0.76, 0.02)	(0.71, 0.03)	(0.79, 0.02)	(0.78, 0.02)	(0.80, 0.02)
Fort	(0.43, 0.04)	(0.39, 0.04)	(0.44, 0.04)	(0.44, 0.04)	(0.41, 0.05)
Utilitaire	(0.85, 0.01)	(0.82, 0.02)	(0.85, 0.01)	(0.83, 0.02)	(0.85, 0.01)
Faible	(0.57, 0.03)	(0.51, 0.04)	(0.58, 0.03)	(0.58, 0.03)	(0.61, 0.04)

TABLE 6.4 – Résultats pour WRC-UCB et WRC-Bandit-UCB

sont atteints avec la stratégie (et type de biais UCB) correspondante à chaque colonne. La colonne étiquettée *Total* représente quant à elle le nombre de stratégie dont les meilleurs

	Superadditive	Classic	Coalitions
	S		
Egalitaire	$(0.88, 0.01)$	$(0.82, 0.01)$	$(0.84, 0.01)$
Égocentrique	$(0.74, 0.03)$	$(0.77, 0.02)$	$(0.81, 0.02)$
Nash	$(0.83, 0.02)$	$(0.79, 0.02)$	$(0.81, 0.02)$
Pareto	$(0.79, 0.02)$	$(0.77, 0.02)$	$(0.79, 0.02)$
Fort	$(0.45, 0.05)$	$(0.46, 0.04)$	$(0.47, 0.05)$
Utilitaire	$(0.85, 0.01)$	$(0.80, 0.02)$	$(0.82, 0.01)$
Faible	$(0.58, 0.03)$	$(0.66, 0.03)$	$(0.66, 0.03)$

TABLE 6.5 – Résultats pour WRC-Superadditive-UCB, WRC-Classic et WRC-Coalitions

résultats correspondent au type de concession de chaque ligne. Le code couleur identifie les meilleurs résultats : en jaune, ce couple de type et stratégie de concession obtient les meilleurs résultats pour cette stratégie, en vert les meilleurs résultats pour ce type de concession, et en bleu, les deux. Les nuances de rouge permettent d'identifier quelles stratégies et types ont le plus de meilleurs résultats, plus le rouge est foncé, plus la stratégie ou le type obtient de bons résultats.

6.3.3 Analyse des résultats

Les premiers résultats pouvant être extraits du tableau sont que le meilleur type de concession est le type *égalitaire*, qui donne les meilleurs résultats pour 15 des 18 stratégies. Le type *utilitaire*, bien que moins performant que le précédent, atteint néanmoins les meilleurs résultats avec 7 des 18 stratégies. Concernant les stratégies, les deux plus performantes se trouvent être WRC-Additional-UCB avec le biais UCB moyen, et WRC-Coalitions, c'est-à-dire l'adaptation sans exploration qui avait été faite pour le cadre distribué. Derrières elles se trouvent WRC-UCB avec le biais UCB moyen, et WRC-Classic, à nouveau l'adaptation sans exploration provenant du cadre distribué. Ces résultats doivent cependant être observés de plus près pour comprendre comment des stratégies sans exploration performent ici. En effet, si nous prenons WRC-Coalitions, nous pouvons voir que celle-ci est plus perfomante que les autres avec le type *égocentrique*, et la plus performante avec WRC-Classic sur les types *fort* et *faible*. Ces trois types, et surtout les deux derniers, sont ceux qui produisent les moyennes de ratio à l'optimal-protocole les plus basses. Nous pouvons donc conclure sur ces types sous-performants qu'une approche gloutonne (et donc sans exploration) est la plus adaptée. En revanche, pour les autres types, les deux

stratégies WRC-Additional-UCB et WRC-UCB sont les plus indiquées, notamment avec l'utilisation conjointe du biais d'exploration UCB moyen, cependant, si nous observons plus attentivement les résultats de ces deux stratégies, nous pouvons souligner le fait que là où WRC-UCB performe (types *égalitaire*, *Nash* et *utilitaire*), WRC-Additional-UCB est toujours au moins aussi bonne. Cette dernière stratégie, avec son biais UCB moyen, est donc la plus adaptée dans ce cadre stochastique.

6.4 Expérimentations et résultats : décentralisé stochastique contre distribué stochastique

Les expérimentations concernant la décentralisation du protocole distribué stochastique peuvent désormais être menées. Il s'agit donc d'utiliser le protocole adapté à un cadre stochastique décentralisé sur des jeux de coalitions stochastiques répétés. Nous pourrons ainsi analyser l'effet de la décentralisation du protocole sur la moyenne du ratio à l'optimal-protocole, mais également si les mêmes types, stratégies et biais UCB efficaces dans le cadre distribué le restent dans ce cadre décentralisé.

6.4.1 Paramètres des expérimentations

Le tableau 6.6 récapitule les différentes configurations entre les stratégies adaptées au cadre décentralisé avec les divers biais d'exploration. À noter la présence de deux stratégies du protocole définies dans le cadre décentralisé (voir section 5.2.2.1) : *WRC-Coalitions-Dec* et *WRC-Classic-Dec*. L'ajout de ces stratégies nous permet d'avoir une comparaison à des stratégies n'intégrant pas d'exploration.

6.4.2 Tableaux de résultats

Nous présentons ci-après les résultats des expérimentations du protocole dans un cadre décentralisé et stochastique sous forme de tableaux, qui montrent les moyennes et variance du ratio à l'optimal-protocole sur l'ensemble des pas de temps des 100 jeux pour chaque paramétrage possible du protocole, à savoir le type de concession pour les lignes, la stratégie de concession pour les colonnes, et si tel est le cas, le type de biais d'exploration UCB décentralisé utilisé (montré par les sous-colonnes des stratégies). Nous choisissons de montrer uniquement le ratio à l'optimal-protocole car elle est la plus repré-

Stratégie	UCB-C-Dec	UCB-I-Dec	UCB-M-Dec	UCB-S-Dec	NO-UCB
WRC-Add.-Dec	✓	✓	✓	✓	✗
WRC-Prop.-Dec	✓	✓	✓	✓	✗
WRC-UCB.-Dec	✓	✓	✓	✓	✗
WRC-Frac.-Dec	✗	✗	✗	✓	✗
WRC-Bandit.-Dec	✗	✗	✗	✓	✗
WRC-Sup.-Dec	✗	✗	✗	✓	✗
WRC-Cl-UCB.-Dec	✗	✗	✗	✓	✗
WRC-Class.-Dec	✗	✗	✗	✗	✓
WRC-Coal.-Dec	✗	✗	✗	✗	✓

TABLE 6.6 – Récapitulatif des stratégies et biais paramétrant le protocole

sentative, un ratio de 1 étant le meilleur résultat possible. Ainsi, le tableau 6.7 montre les résultats pour tout type de concessions et les stratégies WRC-Additional-UCB-Dec et WRC-Classic-UCB-Dec, le tableau 6.8 pour les stratégies WRC-Proportionalized-UCB-Dec et WRC-Fractional-UCB-Dec, le tableau 6.9 pour les stratégies WRC-UCB-Dec et WRC-Bandit-UCB-Dec, et enfin le tableau 6.10 pour les stratégies WRC-Superadditive-UCB-Dec, WRC-Classic-Dec et WRC-Coalitions-Dec.

	Additionnal-Dec				Classic_UCB-Dec
	C-Dec	I-Dec	M-Dec	S-Dec	S-Dec
Egalitaire	(0.53, 0.03)	(0.53, 0.02)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.02)
Égocentrique	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.54, 0.03)	(0.53, 0.03)
Nash	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Pareto	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Fort	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Utilitaire	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Faible	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)

TABLE 6.7 – Résultats pour WRC-Additional-UCB-Dec et WRC-Classic-UCB-Dec

La figure 6.2 en page 168 donne le tableau des résultats complet. La ligne étiquettée *Total* de ce tableau représente le nombre de type de concession dont les meilleurs résultats sont atteints avec la stratégie (et type de biais UCB) correspondante à chaque colonne. La colonne étiquettée *Total* représente quant à elle le nombre de stratégie dont les meilleurs résultats correspondent au type de concession de chaque ligne. Le code couleur identifie les meilleurs résultats : en jaune, ce couple de type et stratégie de concession obtient

	Propotionalized-Dec				Fractional-Dec
	C-Dec	I-Dec	M-Dec	S-Dec	S-Dec
Egalitaire	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Égocentrique	(0.53, 0.03)	(0.53, 0.03)	(0.55, 0.03)	(0.52, 0.03)	(0.58, 0.03)
Nash	(0.53, 0.02)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)
Pareto	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.54, 0.03)
Fort	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.55, 0.03)
Utilitaire	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Faible	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)

TABLE 6.8 – Résultats pour WRC-Proportionalized-UCB-Dec et WRC-Fractional-UCB-Dec

	UCB-Dec				Bandit-Dec
	C-Dec	I-Dec	M-Dec	S-Dec	S-Dec
Egalitaire	(0.53, 0.02)	(0.53, 0.03)	(0.52, 0.02)	(0.52, 0.03)	(0.53, 0.03)
Égocentrique	(0.52, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.55, 0.03)
Nash	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Pareto	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.02)	(0.54, 0.03)	(0.54, 0.03)
Fort	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)
Utilitaire	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)
Faible	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)

TABLE 6.9 – Résultats pour WRC-UCB-Dec et WRC-Bandit-UCB-Dec

	Superadditive-Dec	Classic-Dec	Coalitions-Dec
	S-Dec		
Egalitaire	(0.53, 0.03)	(0.53, 0.02)	(0.53, 0.03)
Égocentrique	(0.54, 0.03)	(0.59, 0.03)	(0.59, 0.03)
Nash	(0.53, 0.03)	(0.54, 0.03)	(0.53, 0.03)
Pareto	(0.53, 0.03)	(0.55, 0.03)	(0.55, 0.03)
Fort	(0.54, 0.03)	(0.54, 0.03)	(0.55, 0.03)
Utilitaire	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)
Faible	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)

TABLE 6.10 – Résultats pour WRC-Superadditive-UCB-Dec, WRC-Classic-Dec et WRC-Coalitions-Dec

les meilleurs résultats pour cette stratégie, en vert les meilleurs résultats pour ce type de concession, et en bleu, les deux. Les nuances de rouge permettent d'identifier quelles stratégies et types ont le plus de meilleurs résultats, plus le rouge est foncé, plus la stratégie ou le type obtient de bons résultats. Attention, les noms utilisés dans ce tableau ne précisent pas que ce sont les versions décentralisées, pour une raison de place et de lisibilité.

6.4.3 Analyse des résultats

Un premier élément à souligner dans ces résultats est la proximité des valeurs pour l'ensemble des configurations possibles, avec un ratio moyen à l'optimal-protocole minimal de 0.52, pour une valeur maximale de 0.59. De plus, ces valeurs sont sujettes à davantage de variance que dans le cadre stochastique distribué, avec une variance de 0.03 sur presque l'ensemble des résultats. La proximité des résultats rend l'analyse par types et stratégies plus délicate, mais il en ressort néanmoins que les types *égocentrique* et *faible*, qui sont les moins bon du cadre distribué déterministe, sont efficaces dans ce cadre décentralisé stochastique. De plus, au niveau des stratégies, les valeurs de ratio moyen seront plus explicites. Tandis que la plupart des stratégies ont un ratio moyen de 0.53 avec la plupart des types, ce n'est pas le cas des stratégies WRC-Classic-Dec, WRC-Coalitions-Dec, et dans une moindre mesure, WRC-Fractional-UCB-Dec. Les deux premières sont notamment celles qui produisent un ratio moyen de 0.59, toutes deux avec le type *égocentrique*, et 0.58 pour la stratégie WRC-Fractional-UCB-Dec sur ce même type. Ces trois stratégies produisent également quelques ratio de 0.54 ou 0.55 avec la plupart des types. Nous avons donc au final deux stratégies sans exploration, et une avec exploration, qui sont plus efficaces que les autres, notamment avec les deux types de concession les moins efficaces dans le cadre distribué déterministe.

6.5 Conclusion

Nous avons proposé dans ce chapitre deux nouvelles adaptations d'un protocole de concessions monotones pour la formation de coalitions, ici pour des cadres stochastiques distribué et décentralisé. Cela a consisté dans un premier temps en la proposition de nouvelles stratégies, pour lesquelles nous avons adapté le biais d'exploration coalitionnel aux propositions du protocole de quatre manières différentes. Ces premiers travaux ont per-

mis d'utiliser le protocole pour résoudre des jeux de coalitions stochastiques répétés, pour lesquels nous pouvons conclure que l'utilisation du type *égalitaire* reste une des meilleures options, comme c'était le cas dans le cadre distribué déterministe. En revanche, les types *Pareto* et *fort* qui étaient tout aussi bons dans ce cadre précédent, ne le sont plus dans ce cadre distribué stochastique, contrairement au type *utilitaire* qui devient le deuxième plus efficace. Concernant les stratégies et biais d'exploration, il ressort que les stratégies WRC-Additional-UCB et WRC-UCB sont efficaces avec le biais d'exploration moyen, ainsi que les stratégies gloutonnes (donc sans exploration) WRC-Coalitions et WRC-Classic que nous avions définies pour le cadre distribué déterministe. Ces deux dernières stratégies sont notamment efficaces avec les types qui le sont le moins. Un constat similaire peut être fait pour la version décentralisée de ce protocole stochastique, où les deux stratégies gloutonnes WRC-Coalitions-Dec et WRC-Classic-Dec donnent les meilleures moyennes de ratio à l'optimal-protocole, tout type confondu. Parmi ces derniers, les types *égocentrique* et *faible*, qui étaient les plus mauvais du cadre distribué déterministe, deviennent les meilleurs, ce qui nous permet de conclure qu'une approche gloutonne et égocentrique est plus adaptée dans un cadre où l'information est manquante.

	additional				proportionalized				UCB				classic_UCB	fractional	structural	superadd.	classic	coalitions	Total
	C	I	M	S	C	I	M	S	C	I	M	S							
Egalitarian	(0.88, 0.01)	(0.86, 0.02)	(0.87, 0.01)	(0.88, 0.01)	(0.75, 0.03)	(0.64, 0.03)	(0.86, 0.02)	(0.66, 0.03)	(0.86, 0.01)	(0.80, 0.02)	(0.86, 0.01)	(0.87, 0.01)	(0.86, 0.01)	(0.82, 0.02)	(0.87, 0.01)	(0.88, 0.01)	(0.82, 0.01)	(0.84, 0.01)	15
Egocentric	(0.77, 0.02)	(0.72, 0.03)	(0.76, 0.02)	(0.75, 0.03)	(0.72, 0.02)	(0.68, 0.03)	(0.76, 0.03)	(0.69, 0.03)	(0.76, 0.02)	(0.73, 0.02)	(0.77, 0.02)	(0.75, 0.03)	(0.74, 0.02)	(0.71, 0.02)	(0.76, 0.03)	(0.74, 0.03)	(0.77, 0.02)	(0.81, 0.02)	3
Nash	(0.83, 0.01)	(0.81, 0.02)	(0.84, 0.01)	(0.83, 0.02)	(0.72, 0.03)	(0.64, 0.04)	(0.83, 0.02)	(0.66, 0.04)	(0.83, 0.01)	(0.79, 0.02)	(0.84, 0.01)	(0.81, 0.02)	(0.80, 0.02)	(0.78, 0.02)	(0.83, 0.02)	(0.83, 0.02)	(0.79, 0.02)	(0.81, 0.02)	2
Pareto	(0.79, 0.02)	(0.79, 0.02)	(0.81, 0.02)	(0.79, 0.02)	(0.65, 0.04)	(0.55, 0.04)	(0.81, 0.02)	(0.56, 0.04)	(0.76, 0.02)	(0.71, 0.03)	(0.79, 0.02)	(0.78, 0.02)	(0.77, 0.02)	(0.74, 0.02)	(0.80, 0.02)	(0.79, 0.02)	(0.77, 0.02)	(0.79, 0.02)	2
Strong	(0.45, 0.04)	(0.45, 0.05)	(0.45, 0.04)	(0.45, 0.04)	(0.37, 0.04)	(0.30, 0.02)	(0.45, 0.05)	(0.34, 0.03)	(0.43, 0.04)	(0.39, 0.04)	(0.44, 0.04)	(0.44, 0.04)	(0.43, 0.04)	(0.45, 0.04)	(0.41, 0.05)	(0.45, 0.05)	(0.46, 0.04)	(0.47, 0.05)	2
Utilitarian	(0.86, 0.01)	(0.84, 0.02)	(0.86, 0.01)	(0.85, 0.01)	(0.74, 0.03)	(0.64, 0.04)	(0.83, 0.02)	(0.68, 0.04)	(0.85, 0.01)	(0.82, 0.02)	(0.85, 0.01)	(0.83, 0.02)	(0.83, 0.01)	(0.81, 0.02)	(0.85, 0.01)	(0.85, 0.01)	(0.80, 0.02)	(0.82, 0.01)	7
Weak	(0.58, 0.03)	(0.57, 0.04)	(0.60, 0.03)	(0.59, 0.03)	(0.52, 0.04)	(0.42, 0.04)	(0.61, 0.03)	(0.45, 0.04)	(0.57, 0.03)	(0.51, 0.04)	(0.58, 0.03)	(0.58, 0.03)	(0.54, 0.03)	(0.53, 0.03)	(0.61, 0.04)	(0.58, 0.03)	(0.66, 0.03)	(0.66, 0.03)	2
Total	2	1	4	1	2	1	2	2	2	1	3	1	1	1	1	1	3	4	

Vertical	+++
Horizontal	++
Les deux	+

FIGURE 6.1 – Tableau des moyennes et variances du ratio à l'optimal protocole (versions distribuées)

	additional			proportionalized			UCB			classic_UCB	fractional	bandit	superadd.	classic	coalitions	Total
	C	I	M	S	C	I	M	S	C	I	M	S								
Egalitarian	(0.53, 0.03)	(0.53, 0.02)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.02)	(0.53, 0.03)	(0.52, 0.02)	(0.52, 0.03)	(0.53, 0.02)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	8
Egocentric	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.54, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.55, 0.03)	(0.52, 0.03)	(0.52, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.53, 0.03)	(0.58, 0.03)	(0.55, 0.03)	(0.54, 0.03)	(0.59, 0.03)	(0.59, 0.03)	(0.59, 0.03)	9
Nash	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.02)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.53, 0.03)	8
Pareto	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.02)	(0.54, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.53, 0.03)	(0.55, 0.03)	(0.55, 0.03)	(0.55, 0.03)	8
Strong	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.55, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.55, 0.03)	7
Utilitarian	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.52, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.54, 0.03)	4
Weak	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.53, 0.03)	(0.54, 0.03)	(0.54, 0.03)	(0.54, 0.03)	9
Total	5	1	5	1	1	6	1	5	1	5	1	2	1	4	2	2	5	5		

Vertical
Horizontal
Les deux

+++
++
+

FIGURE 6.2 – Tableau des moyennes et variances du ratio à l'optimal protocole (versions décentralisées)

CONCLUSION ET PERSPECTIVES

Sommaire

Problématiques

De nombreuses applications réelles peuvent être modélisées grâce aux systèmes multi-agents, et lorsque que les agents doivent coopérer pour réaliser certaines tâches, nous pouvons modéliser ce problème par des jeux de coalitions. Cependant, certains cadres applicatifs nécessitent la levée d'hypothèses communément faites dans la formation de coalitions, en raison de leur incompatibilité avec ces applications réelles. Ces hypothèses sont les suivantes.

1. Déterminisme des utilités des coalitions,
2. Connaissance *a priori* des utilités par les agents,
3. Centralisation du problème de formation de coalitions.

Nous avons dans un premier temps abordé la levée des hypothèses de déterminisme et de connaissances *a priori* des utilités, et proposé pour cela un modèle pour la formation de coalitions stochastique répétée qui s'abstrait de ces hypothèses ainsi que deux concepts de solutions fondés sur une notion d'équilibre exploration-exploitation bien connue du domaine de l'apprentissage par renforcement, nommés δ-cœur et γ-cœur.

Afin d'aborder la problématique concernant la décentralisation du problème de formation de coalitions, nous avons décidé d'adapter un protocole de concessions monotones pour la négociation multilatérale entre agents utilitaristes, proposé par Ulle Endriss. En effet, ce protocole est exempt d'hypothèse sur la structure du système et ne contient aucune entité centrale telle qu'un commissaire-priseur, qui est un élément simulant une centralisation du système. Nous avons donc proposé dans un premier temps une adaptation de ce protocole à la formation de coalitions dans un cadre distribué, avant d'en

proposer une version dans un cadre décentralisé. Enfin, nous avons travaillé à l'unification des modèles que nous avons proposé pour la levée des différentes hypothèses, en proposant tout d'abord une adaptation du protocole distribué au cadre stochastique et répété, puis une adaptation dans ce même cadre du protocole décentralisé. Ainsi, nous proposons un modèle unique où les trois hypothèses dont nous souhaitions nous abstraire sont levées. Le traitement des hypothèses dans ce mémoire est résumé par le tableau 6.11.

Hypothèse	Chapitres
Déterminisme des utilités	3, 4 & 6
Connaissances *a priori*	3, 4 & 6
Centralisation	5 & 6

TABLE 6.11 – Hypothèses abordées dans les différents chapitres

Contributions

Dans les chapitres 3 et 4, nous nous sommes donc intéressés à la levée des hypothèses de déterminisme et de connaissance *a priori* des utilités des coalitions en proposant un modèle de jeux de coalitions stochastiques répétés ainsi que deux concepts de solutions fondés sur l'équilibre exploration-exploitation : γ-cœur et δ-cœur, que nous avons mis en concurrence avec un concept glouton, à savoir ϵ-cœur. Nous avons ainsi montré que le concept γ-cœur permet un meilleur apprentissage de la fonction caractéristique que les autres concepts, mais ne produit cependant pas un meilleur regret. Le concept δ-cœur quant à lui permet un apprentissage semblable à celui de la méthode gloutonne, ϵ-cœur, mais ce dernier concept génère en général un regret aussi bon ou meilleur que δ-cœur. Ainsi, un meilleur apprentissage de la fonction caractéristique n'améliore pas nécessairement les résultats, car les agents n'ont pas besoin de connaître les valeurs exactes des coalitions tant qu'ils arrivent à créer un ordre entre les différentes coalitions.

Dans le chapitre 5, nous nous sommes intéressés à la décentralisation du problème de formation de coalitions. Pour cela, nous avons dans un premier temps adapté un protocole de concessions monotones pour la négociation multilatérale au problème de formation de coalitions, pour lequel nous avons redéfini une stratégie afin d'y intégrer la notion de coalitions, à partir de laquelle nous avons donc proposé les nouvelles stratégies *WRC-Surplus*, *WRC-Coalitions* et *WRC-Classic*. Nous avons montré que la stratégie *WRC-*

Coalitions était la plus performante, notamment avec trois types de concession, à savoir *égalitaire*, *fort* et *Pareto*. Dans un second temps, nous avons restreint les connaissances des agents uniquement aux coalitions auxquelles ils peuvent appartenir. Ainsi, les agents doivent décider d'une structure de coalitions de façon décentralisée, mais à partir d'une connaissance restreinte. Pour cela, nous avons proposé de nouvelles stratégies, fondées sur la stratégie *WRC* originale (à savoir *WRC-Classic*, et sur *WRC-Coalitions*, qui était la meilleure dans le cadre décentralisé. Les types de concession efficaces dans un cadre décentralisé ne sont pas les mêmes, avec un type *égocentrique* plus performant, ainsi que la stratégie *WRC-Classic-Dec* qui ne prend pas en compte la notion de coalitions. Cependant, les pertes d'optimalité sont bien plus importantes dans le cadre décentralisé. Nous avons ainsi pu mettre en lumière que, comme attendu, le protocole distribué est plus performant que le protocole décentralisé, en revanche, les stratégies et types les moins performantes dans le cadre distribué deviennent les meilleures dans le cadre décentralisé.

Le chapitre 6 a été consacré à l'extension du protocole adapté à la formation de coalitions à un cadre stochastique et répété afin d'aborder conjointement les trois hypothèses que nous souhaitions lever. Dans un premier temps, cela a consisté en l'adaptation aux jeux de coalitions stochastiques répétés du protocole distribué. Nous avons dans un premier temps adapté un biais d'exploration coalitionnel défini pour la première fois dans le chapitre 3 afin de pouvoir l'utiliser dans les stratégies, avec différentes sémantiques (exploration collective, individuelle, moyenne ou structurelle). Avec ces biais, nous avons donc proposé de nouvelles adaptations des stratégies *WRC-Coalitions* et *WRC-Classic*. Nous avons montré que dans un cadre stochastique distribué, le type *égalitaire* se distingue, qu'un biais d'exploration de type moyen est plus performant, tout comme la stratégie *WRC-Additional-UCB*. Sur les types les moins performants, les stratégies gloutonnes *WRC-Classic* et *WRC-Coalitions* restent les meilleures, en cohérence avec les résultats de chapitre 4. Nous avons ensuite fusionné ce dernier protocole avec le protocole adapté au cadre décentralisé défini dans le chapitre 5, et par conséquent adapté les stratégies stochastiques au cadre décentralisé. Lorsque nous passons dans ce cadre décentralisé, les types les plus performants changent, à la faveur des types plus individuels comme *égocentrique* et *faible*. De plus, les stratégies gloutonnes *WRC-Classic-Dec* et *WRC-Coalitions-Dec* surpassent les autres.

Le tableau 6.12 récapitule les meilleurs paramètres des protocoles à utiliser dans les différents cadres expérimentés. Il est important de noter que l'ensemble de nos modèles, nous ne faisons pas d'hypothèses fortes sur les jeux de coalitions que nous résolvons.

	Distribué	Décentralisé
Déterministe	(WRC-Coalitions, égalitaire), (WRC-Coalitions, Pareto), (WRC-Coalitions, fort)	(WRC-Classic-Dec, égocentrique), (WRC-Coalitions-Dec, égocentrique)
Stochastique	(WRC-Add.-UCB-M, égalitaire), (WRC-UCB-M, égalitaire)	(WRC-Classic-Dec, égocentrique), (WRC-Coalitions-Dec, égocentrique)

TABLE 6.12 – Paramètres les plus performants selon le cadre

Les travaux menés dans le cadre de ce mémoire ont fait l'objet de plusieurs publications. La levée des hypothèses de déterminisme et connaissances *a priori* des utilités a été traitée dans les articles [Guéneron et Bonnet, 2020, Guéneron et Bonnet, 2021a]. L'adaptation du protocole au cadre de la formation de coalitions a été traitée dans l'article [Guéneron et Bonnet, 2022]. Enfin, un article faisant un état de l'art sur la diversité de la formation de coalitions a également été produit [Guéneron et Bonnet, 2021b].

Perspectives

Les travaux que nous avons menés nous amènent à nous poser de nouvelles questions. Nos perspectives suivent trois axes différents. Un premier concerne directement les approches utilisées dans les différents modèles, notamment en ce qui concerne les concepts de solutions et l'apprentissage. Un deuxième axe regroupe les perspectives concernant la distribution des gains entre les agents dans les protocoles proposés dans les chapitres 5 et 6. Enfin, le dernier axe concerne des perspectives plus générales quant à la gestion des pannes dans nos modèles ainsi que leur robustesse.

Décision multi-critères et apprentissage

Nous avons, dans les chapitres 3 et 4, proposé et expérimenté des concepts de solutions fondés sur un principe d'exploration-exploitation en intégrant à un concept de solutions pré-existant un terme d'exploration. La force du terme d'exploration étant délicate à doser dans l'équilibre exploration-exploitation, nous pouvons nous demander si une autre approche est envisageable. Le domaine de la décision multi-objectifs est alors une piste intéressante. Cela permet notamment l'optimisation de fonctions de récompense fondées sur de multiples critères. Des travaux utilisant l'optimisation multi-objectifs dans la for-

mation de coalitions ont déjà été menés pour l'intégration d'un critère de confiance [Cho *et al.*, 2013]. Il semble donc pertinent d'imaginer aborder la formation de coalitions stochastique par le prisme de la décision multi-objectifs, où l'exploration serait un critère au même titre que les récompenses.

Dans les travaux présentés dans les chapitres 5 et 6, nous avons utilisé un algorithme d'apprentissage tabulaire, c'est-à-dire que nous apprenons les valeurs des coalitions une par une, tandis que dans les chapitres 3 et 4, nous avions deux modèles d'apprentissage différents : un tabulaire, et un par inférence. Une première perspective concernant l'apprentissage est donc d'utiliser également un réseau de neurones conjointement aux protocoles adaptés que nous avons proposé. Une deuxième perspective ici serait un modèle d'apprentissage utilisant une structure plus compacte que la méthode tabulaire, mais également avec un fonctionnement plus explicable que les réseaux de neurones. Une piste pour cela serait un apprentissage sur une structure compacte de fonctions caractéristiques comme les réseaux de contributions marginales – appelés *MC-nets* dans la littérature [Elkind *et al.*, 2009]. Toutefois, un verrou scientifique concernant ces réseaux est la représentation de fonctions caractéristiques dont la structure ne possède pas de propriétés fortes comme la superaddivitité.

Règles de distribution

En raison des divers résultats des chapitres 5 et 6, deux perspectives concernent les protocoles adaptés proposés pour la levée des diverses hypothèses. Dans ceux-ci, les agents construisent des propositions en s'appuyant sur une règle de distribution particulière, incluant d'une part un gage minimum, et de l'autre part une négociation sur le surplus (c'est-à-dire la part d'utilité restante dans une coalition après le gage minimum pour chaque agent retiré) grâce au protocole d'Ulle Endriss, ce qui correspond à une distribution égalitaire de ce surplus. Il est donc envisageable de proposer d'autres règles de distribution, afin d'explorer davantage l'espace des solutions possibles, permettant potentiellement d'atteindre des solutions plus proches de l'optimal. Une règle fondée sur la contribution marginale des agents est une piste.

Une deuxième perspective, liée directement à la précédente, est d'envisager que les agents puissent négocier sur ces règles de distribution. Nous pouvons par exemple supposer qu'une coalition puisse ne pas être stable avec une première règle de distribution, mais qu'elle le soit avec une autre. Cependant, cela dépend également de la structure de coalitions dans lesquelles ces règles sont appliquées. Une négociation concernant la règle

de distribution utilisée est donc une piste envisageable.

Robustesse et modèles de pannes

Les jeux de coalitions stochastiques répétés que nous proposons intègrent un modèle stochastique fondé sur des gaussiennes, cela nous permettant de modéliser des utilités stochastiques dépendant d'aléas ponctuels environnementaux ou propres aux agents. Toutefois, ce type de modèle ne permet pas de modéliser correctement une notion de panne. En effet, nous pouvons imaginer que dans une application réelle de système multi-agents (par exemple le port maritime), il peut arriver qu'un agent tombe en panne. Une piste à envisager est donc l'utilisation d'autres modèles stochastiques permettant de modéliser de tels cas. Si une panne se produit avant de former les coalitions, il suffit de ne pas prendre l'agent en compte, cependant, si cela se passe après la formation de coalitions, le déroulement du jeu doit y être robuste, c'est-à-dire que la coopération des autres agents continue. Pour un concept de solutions, cela peut être inspiré de travaux intégrant une probabilité de succès des tâches que les agents ont à exécuter, probabilité qui changerait en fonction des pannes des agents. Concernant le protocole, lorsqu'un agent tombe en panne dans une coalition, les agents peuvent simplement considérer qu'il est absent et pas conséquent former la même coalition mais sans lui, avec l'utilité correspondante décrite par la fonction caractéristique.

Cette question de robustesse peut être étendue aux protocoles que nous avons proposé, mais pas seulement dans un cadre stochastique. Que se passerait-il en cas d'agents incompétents qui ne produisent aucune utilité dans toutes les coalitions auxquelles ils appartiennent ? En effet, un agent ne produisant aucun gain (appelé *dummy player* dans la littérature) peut faire néanmoins partie de coalitions ayant un surplus positif, coalitions qu'il proposera. Or, la plupart des autres agents risquent de le proposer seul, et donc n'ayant aucun gain, il risque alors de ne pas concéder, car sinon il perdrait tout son gain. D'autre part, est-ce que les protocoles proposés sont robustes à la manipulation ? Par exemple quel impact aurait un agent qui déciderait de communiquer une fausse valeur d'utilité pour sa coalition singleton dans le cadre décentralisé ? Ces questions sont importantes pour l'utilisation de ces modèles dans des applications réelles afin que des comportements indésirables ne perturbent pas le système. Une piste envisageable est d'utiliser les notions de confiance et de réputation [Sabater et Sierra, 2005], afin de potentiellement exclure des agents indésirables du système.

BIBLIOGRAPHIE

[Aadithya *et al.*, 2011] Aadithya, K. V., Michalak, T. P. et Jennings, N. R. (2011). Representation of coalitional games with algebraic decision diagrams. *In 10th AAMAS*, pages 1121–1122. [p. 38]

[Adam *et al.*, 2011] Adam, E., Hette, G., Estivie, S., Melki, A. et Mandiau, R. (2011). Agents tasks reallocation for collaborative urban supply chain management. *In International Conference on Industrial Applications of Holonic and Multi-Agent Systems*, pages 215–224. Springer. [p. 5]

[Agrawal, 1995] Agrawal, R. (1995). Sample mean based index policies by o (log n) regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078. [p. 52], [p. 53], [p. 68]

[Airiau et Sen, 2010] Airiau, S. et Sen, S. (2010). On the stability of an optimal coalition structure. *In 19th ECAI*, pages 203–208. [p. 23]

[Anantharam *et al.*, 1987] Anantharam, V., Varaiya, P. et Walrand, J. (1987). Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-part i : Iid rewards. *IEEE Transactions on Automatic Control*, 32(11):968–976. [p. 55]

[Anshelevich *et al.*, 2008] Anshelevich, E., Dasgupta, A., Kleinberg, J., Tardos, É., Wexler, T. et Roughgarden, T. (2008). The price of stability for network design with fair cost allocation. *SIAM Journal on Computing*, 38(4):1602–1623. [p. 26], [p. 128], [p. 129]

[Apt et Witzel, 2009] Apt, K. R. et Witzel, A. (2009). A generic approach to coalition formation. *Int. Game Theory Rev.*, 11(3):347–367. [p. 29], [p. 38]

[Aubin, 1981] Aubin, J.-P. (1981). Cooperative fuzzy games. *Math. Oper. Res.*, 6:1–13. [p. 28], [p. 29], [p. 38]

[Audibert *et al.*, 2011] Audibert, J.-Y., Bubeck, S. et Lugosi, G. (2011). Minimax policies for combinatorial prediction games. *In 24th COLT*, pages 107–132. [p. 56]

[Auer *et al.*, 2002] Auer, P., Cesa-Bianchi, N. et Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256. [p. 52], [p. 53], [p. 58]

[Auer *et al.*, 1995] Auer, P., Cesa-Bianchi, N., Freund, Y. et Schapire, R. E. (1995). Gambling in a rigged casino : The adversarial multi-armed bandit problem. *In 36th FOCS*, pages 322–331. IEEE. [p. 54]

[Axelrod et Hamilton, 1981] Axelrod, R. et Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489):1390–1396. [p. 14], [p. 48]

[Aziz et De Keijzer, 2011] Aziz, H. et De Keijzer, B. (2011). Complexity of coalition structure generation. *In 11th AAMAS*, pages 191–198. [p. 29]

[Bachrach *et al.*, 2010] Bachrach, Y., Meir, R., Jung, K. et Kohli, P. (2010). Coalitional structure generation in skill games. *In 24th AAAI*. [p. 29]

[Bachrach et Rosenschein, 2008] Bachrach, Y. et Rosenschein, J. S. (2008). Coalitional skill games. *In 7th AAMAS*, pages 1023–1030. [p. 28], [p. 29], [p. 38]

[Banzhaf III, 1964] Banzhaf III, J. F. (1964). Weighted voting doesn't work : A mathematical analysis. *Rutgers Univ. Law Rev.*, 19:317. [p. 19], [p. 38]

[Bell, 1938] Bell, E. T. (1938). The iterated exponential integers. *Annals of Mathematics*, pages 539–557. [p. 33], [p. 128]

[Bistaffa *et al.*, 2017] Bistaffa, F., Farinelli, A., Cerquides, J., Rodríguez-Aguilar, J. et Ramchurn, S. D. (2017). Algorithms for graph-constrained coalition formation in the real world. *ACM Trans. Intell. Syst. Technol.*, 8(4):60 :1–60 :24. [p. 34], [p. 35], [p. 36], [p. 38]

[Blankenburg *et al.*, 2005] Blankenburg, B., Dash, R. K., Ramchurn, S. D., Klusch, M. et Jennings, N. R. (2005). Trusted kernel-based coalition formation. *In 4th AAMAS*, pages 989–996. [p. 38], [p. 44], [p. 48]

[Blankenburg *et al.*, 2003] Blankenburg, B., Klush, M. et Shehory, O. (2003). Fuzzy kernel-stable coalitions between rational agents. *In 2nd AAMAS*, pages 9–16. [p. 38], [p. 46]

[Bonnet et Tessier, 2007] Bonnet, G. et Tessier, C. (2007). Collaboration among a satellite swarm. *In 6th AAMAS*, pages 1–8. [p. 5]

[Bonzon, 2007] Bonzon, E. (2007). *Modélisation des interactions entre agents rationnels : les jeux booléens*. Thèse de doctorat, Paul Sabatier University, Toulouse, France. [p. 15]

[Borkotokey et Neog, 2014] Borkotokey, S. et Neog, R. (2014). Dynamic resource allocation in fuzzy coalitions : a game theoretic model. *Fuzzy Optim. Decis. Mak.*, 13:211–230. [p. 28], [p. 29], [p. 38]

[Bremer et Lehnhoff, 2017] Bremer, J. et Lehnhoff, S. (2017). Decentralized coalition formation with agent-based combinatorial heuristics. Rapport technique, Ediciones Universidad de Salamanca. [p. 5], [p. 34], [p. 35], [p. 36]

[Brooks, 1986] Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE J. Robotics Autom.*, 2(1):14–23. [p. 9]

[Burkart et Huber, 2021] Burkart, N. et Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *J. Artif. Intell. Res.*, 70:245–317. [p. 65]

[Calvo et Gutiérrez, 2010] Calvo, E. et Gutiérrez, E. (2010). Solidarity in games with a coalition structure. *Math. Soc. Sci.*, 60(3):196–203. [p. 20], [p. 38]

[Chalkiadakis et Boutilier, 2004] Chalkiadakis, G. et Boutilier, C. (2004). Bayesian reinforcement learning for coalition formation under uncertainty. *In 3rd AAMAS*, pages 1090–1097. [p. 28], [p. 38], [p. 46], [p. 47]

[Chalkiadakis et Boutilier, 2008] Chalkiadakis, G. et Boutilier, C. (2008). Sequential decision making in repeated coalition formation under uncertainty. *In 6th AAMAS*, pages 347–354. [p. 28], [p. 38], [p. 46], [p. 49]

[Chalkiadakis *et al.*, 2008] Chalkiadakis, G., Elkind, E., Markakis, E. et Jennings, N. R. (2008). Overlapping coalition formation. *In 4th WINE*, pages 307–321. [p. 30], [p. 38]

[Chalkiadakis *et al.*, 2011] Chalkiadakis, G., Elkind, E. et Wooldridge, M. J. (2011). Computational aspects of cooperative game theory. *Synth. Lect. Artif. Intell. Mach. Learn.*, 5(6):1–168. [p. 15], [p. 16], [p. 18], [p. 22], [p. 27], [p. 31], [p. 33]

[Chalkiadakis *et al.*, 2007] Chalkiadakis, G., Markakis, E. et Boutilier, C. (2007). Coalition formation under uncertainty : Bargaining equilibria and the Bayesian core stability concept. *In 6th AAMAS*, pages 1–8. [p. 38], [p. 46], [p. 47]

[Charnes et Granot, 1973] Charnes, A. et Granot, D. (1973). Prior solutions : Extensions of convex nucleus solutions to chance-constrained games. Rapport technique, Texas Univ. [p. 38], [p. 46], [p. 51], [p. 59], [p. 62]

[Charnes et Granot, 1976] Charnes, A. et Granot, D. (1976). Coalitional and chance-constrained solutions to n-person games. i : The prior satisficing nucleolus. *SIAM J. Appl. Math.*, 31(2):358–367. [p. 38], [p. 46], [p. 59]

[Chen *et al.*, 2013] Chen, W., Wang, Y. et Yuan, Y. (2013). Combinatorial multi-armed bandit : General framework and applications. *In 30th ICML*, pages 151–159. [p. 55]

[Chen *et al.*, 2016] Chen, W., Wang, Y., Yuan, Y. et Wang, Q. (2016). Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *The Journal of Machine Learning Research*, 17(1):1746–1778. [p. 56]

[Cho *et al.*, 2013] Cho, J.-H., Chen, I.-R., Wang, Y., Chan, K. S. et Swami, A. (2013). Multi-objective optimization for trustworthy tactical networks : A survey and insights. Rapport technique, U.S. Army Res. Lab., CISD Aber. Pro. Gro., MD. [p. 173]

[Collier et Llorens, 2018] Collier, M. et Llorens, H. U. (2018). Deep contextual multi-armed bandits. *CoRR*, abs/1807.09809. [p. 57]

[Davis et Maschler, 1965] Davis, M. et Maschler, M. (1965). The kernel of a cooperative game. *Nav. Res. Logist.*, 12(3):223–259. [p. 22], [p. 38]

[Dawson *et al.*, 2009] Dawson, M. R. W., Dupuis, B., Spetch, M. L. et Kelly, D. M. (2009). Simple artificial neural networks that match probability and exploit and explore when confronting a multiarmed bandit. *IEEE Trans. on Neural Networks*, 20(8):1368–1371. [p. 57]

[Deng *et al.*, 2009] Deng, X., Fang, Q. et Sun, X. (2009). Finding nucleolus of flow game. *J. Comb. Optim.*, 18(1):64–86. [p. 30], [p. 38]

[Diago *et al.*, 2016] Diago, N. A., Aknine, S., Shehory, O., Arib, S., Cailliere, R. et Sene, M. (2016). Decentralized and fair multilateral negotiation. *In 28th ICTAI*, pages 149–156. [p. 34], [p. 35]

[Elkind *et al.*, 2009] Elkind, E., Goldberg, L. A., Goldberg, P. W. et Wooldridge, M. J. (2009). A tractable and expressive class of marginal contribution nets and its applications. *Math. Log. Q.*, 55(4):362–376. [p. 173]

[Elkind et Wooldridge, 2009] Elkind, E. et Wooldridge, M. J. (2009). Hedonic coalition nets. *In 8th AAMAS*, pages 417–424. [p. 17]

[Endriss, 2006] Endriss, U. (2006). Monotonic concession protocols for multilateral negotiation. *In 5th AAMAS*, pages 392–399. [p. 108]

[Flood *et al.*, 1950] Flood, M., Dresher, M., Tucker, A. et Device, F. (1950). Prisoner's dilemma : game theory. *Experimental Economics*, 54. [p. 13]

[Gallardo et Jiménez-Losada, 2020] Gallardo, J. M. et Jiménez-Losada, A. (2020). A characterization of the Shapley value for cooperative games with fuzzy characteristic function. *Fuzzy Sets Syst.*, 398:98–111. [p. 38], [p. 46]

[Gaston et desJardins, 2005] Gaston, M. E. et desJardins, M. (2005). Agent-organized networks for dynamic team formation. *In 4th AAMAS*, pages 230–237. [p. 5], [p. 28], [p. 34], [p. 35], [p. 36], [p. 38]

[Gillies, 1959] Gillies, D. B. (1959). Solutions to general non-zero-sum games. *Contributions to the Theory of Games*, 4:47–85. [p. 21], [p. 38]

[Gittins, 1979] Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Jour. of the Royal Stat. Soc.*, 41(2):148–164. [p. 50]

[Glinton *et al.*, 2008] Glinton, R., Scerri, P. et Sycara, K. (2008). Agent-based sensor coalition formation. *In 11th FUSION*, pages 1–7. [p. 5], [p. 34], [p. 35], [p. 36], [p. 38]

[Granot et Granot, 1992] Granot, D. et Granot, F. (1992). On some network flow games. *Math. Oper. Res.*, 17(4):792–841. [p. 30], [p. 38]

[Guéneron et Bonnet, 2020] Guéneron, J. et Bonnet, G. (2020). Un modèle de jeux de coalitions stochastiques répétés. *In 18e RJCIA*, pages 34–41. [p. 172]

[Guéneron et Bonnet, 2021a] Guéneron, J. et Bonnet, G. (2021a). Are exploration-based strategies of interest for repeated stochastic coalitional games ? *In 19th PAAMS*, pages 89–100. [p. 172]

[Guéneron et Bonnet, 2021b] Guéneron, J. et Bonnet, G. (2021b). De la diversité des jeux de coalitions à utilité transférable. *In 29e JFSMA*, pages 149–158. [p. 172]

[Guéneron et Bonnet, 2022] Guéneron, J. et Bonnet, G. (2022). Un protocole de concessions monotones pour la formation distribuée de coalitions. *In 30e JFSMA*, pages 31–40. [p. 172]

[Hansson, 1968] Hansson, B. (1968). Choice structures and preference relations. *Synthese*, pages 443–458. [p. 16]

[Hewitt, 1991] Hewitt, C. (1991). Open information systems semantics for distributed artificial intelligence. *Artif. Intell.*, 47(1-3):79–106. [p. 9]

[Hoos et Stützle, 2004] Hoos, H. H. et Stützle, T. (2004). *Stochastic local search : Foundations and applications*. Elsevier. [p. 34]

[Ieong et Shoham, 2005] Ieong, S. et Shoham, Y. (2005). Marginal contribution nets : A compact representation scheme for coalitional games. *In 6th EC*, pages 193–202. [p. 38]

[Ieong et Shoham, 2008] Ieong, S. et Shoham, Y. (2008). Bayesian coalitional games. *In 23rd AAAI*, pages 95–100. [p. 38], [p. 46], [p. 47], [p. 66]

[Kalai et Zemel, 1982] Kalai, E. et Zemel, E. (1982). Totally balanced games and games of flow. *Math. Oper. Res.*, 7(3):476–478. [p. 30], [p. 38]

[Kern et Paulusma, 2003] Kern, W. et Paulusma, D. (2003). Matching games : the least core and the nucleolus. *Math. Oper. Res.*, 28(2):294–308. [p. 30], [p. 38]

[Khalouzadeh *et al.*, 2010] Khalouzadeh, L., Nematbakhsh, N. et Zamanifar, K. (2010). A decentralized coalition formation algorithm among homogeneous agents. *J. Theor. Appl. Inf. Technol.*, 22(1). [p. 34], [p. 35], [p. 38]

[Kraus *et al.*, 2003] Kraus, S., Shehory, O. et Taase, G. (2003). Coalition formation with uncertain heterogeneous information. *In 2nd AAMAS*, pages 1–8. [p. 28], [p. 38], [p. 44], [p. 45]

[Kurz, 1988] Kurz, M. (1988). Coalitional value. *In* Roth, A., éditeur : *The Shapley value : essays in honor of Lloyd S. Shapley*, pages 155–173. Cambridge University Press. [p. 19]

[Larson et Sandholm, 2000] Larson, K. S. et Sandholm, T. W. (2000). Anytime coalition structure generation : an average case study. *J. Exp. Theor. Artif. Intell.*, 12(1):23–42. [p. 80]

[Mareš, 2001] Mareš, M. (2001). *Fuzzy Cooperative Games*. Springer. [p. 38], [p. 46]

[Marwell et Schmitt, 1972] Marwell, G. et Schmitt, D. R. (1972). Cooperation in a three-person prisoner's dilemma. *Journal of Personality and Social Psychology*, 21(3): 376. [p. 14]

[Maschler *et al.*, 1979] Maschler, M., Peleg, B. et Shapley, L. S. (1979). Geometric properties of the kernel, nucleolus, and related solution concepts. *Math. Oper. Res.*, 4(4):303–338. [p. 22], [p. 23]

[Mauro *et al.*, 2010] Mauro, N. D., Basile, T., Ferilli, S. et Esposito, F. (2010). Coalition structure generation with grasp. *In 14th AIMSA*, pages 111–120. [p. 34]

[Michalak *et al.*, 2010] Michalak, T. P., Rahwan, T., Marciniak, D., Szamotulski, M. et Jennings, N. R. (2010). Computational aspects of extending the Shapley value to coalitional games with externalities. *In 19th ECAI*, pages 197–202. [p. 32], [p. 38]

[Michalak *et al.*, 2009] Michalak, T. P., Rahwan, T., Sroka, J., Dowell, A., Wooldridge, M. J., McBurney, P. J. et Jennings, N. R. (2009). On representing coalitional games with externalities. *In 10th EC*, pages 11–20. [p. 32], [p. 38]

[Mihailescu *et al.*, 2011] Mihailescu, R.-C., Vasirani, M. et Ossowski, S. (2011). Dynamic coalition adaptation for efficient agent-based virtual power plants. *In 9th MATES*, pages 101–112. [p. 34], [p. 35], [p. 36], [p. 38]

[Mochaourab et Jorswieck, 2014] Mochaourab, R. et Jorswieck, E. A. (2014). Coalitional games in MISO interference channels : Epsilon-core and coalition structure stable set. *IEEE Trans. Signal Process.*, 62(24):6507–6520. [p. 22], [p. 38]

[Morge et Nongaillard, 2017] Morge, M. et Nongaillard, A. (2017). Distributed algorithm for egalitarian matching between individuals and activities with additively separable preferences. *In 29th ICTAI*, pages 739–746. [p. 17]

[Morgenstern et Von Neumann, 1953] Morgenstern, O. et Von Neumann, J. (1953). *Theory of games and economic behavior*. Princeton University Press. [p. 7], [p. 10], [p. 11], [p. 14], [p. 15], [p. 18], [p. 38]

[Moulin et Chaib-Draa, 1996] Moulin, B. et Chaib-Draa, B. (1996). *An overview of distributed artificial intelligence*, page 3–55. John Wiley & Sons, Inc. [p. 8]

[Nash, 1951] Nash, J. (1951). Non-cooperative games. *Annals of mathematics*, pages 286–295. [p. 13]

[Nowak et Radzik, 1994] Nowak, A. S. et Radzik, T. (1994). A solidarity value for n-person transferable utility games. *Int. J. Game Theory*, 23(1):43–48. [p. 20], [p. 38]

[Ohta *et al.*, 2006] Ohta, N., Iwasaki, A., Yokoo, M., Maruono, K., Conitzer, V. et Sandholm, T. (2006). A compact representation scheme for coalitional games in open anonymous environments. *In 21st AAAI*, pages 509–514. [p. 28], [p. 38]

[Osborne et Rubinstein, 1994] Osborne, M. J. et Rubinstein, A. (1994). *A course in game theory*. MIT press. [p. 13], [p. 15], [p. 21]

[Othmani-Guibourg *et al.*, 2017] Othmani-Guibourg, M., El Fallah-Seghrouchni, A., Farges, J.-L. et Potop-Butucaru, M. (2017). Multi-agent patrolling in dynamic environments. *In 5th ICA*, pages 72–77. IEEE. [p. 5]

[Paccagnan *et al.*, 2022] Paccagnan, D., Chandan, R. et Marden, J. R. (2022). Utility and mechanism design in multi-agent systems : An overview. *Annu. Rev. Control.*, 53:315–328. [p. 7]

[Poundstone, 1993] Poundstone, W. (1993). *Prisoner's Dilemma/John Von Neumann, game theory and the puzzle of the bomb*. Anchor. [p. 13]

[Rahwan et Jennings, 2008] Rahwan, T. et Jennings, N. R. (2008). An improved dynamic programming algorithm for coalition structure generation. *In 7th AAMAS*, pages 1417–1420. [p. 32]

[Rahwan *et al.*, 2012] Rahwan, T., Michalak, T., Wooldridge, M. J. et Jennings, N. R. (2012). Anytime coalition structure generation in multi-agent systems with positive or negative externalities. *Artif. Intell.*, 186:95–122. [p. 32], [p. 38]

[Rahwan *et al.*, 2015] Rahwan, T., Michalak, T. P., Wooldridge, M. J. et Jennings, N. R. (2015). Coalition structure generation : A survey. *Artif. Intell.*, 229:139–174. [p. 32]

[Rahwan *et al.*, 2009] Rahwan, T., Ramchurn, S. D., Jennings, N. R. et Giovannucci, A. (2009). An anytime algorithm for optimal coalition structure generation. *J. Artif. Intell. Res.*, 34:521–567. [p. 34], [p. 80]

[Rota, 1964] Rota, G.-C. (1964). The number of partitions of a set. *The American Mathematical Monthly*, 71(5):498–504. [p. 128]

[Russell et Norvig, 1995] Russell, S. et Norvig, P. (1995). *Artificial intelligence : a modern approach*. Prentice Hall. [p. 7]

[Sabater et Sierra, 2005] Sabater, J. et Sierra, C. (2005). Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60. [p. 48], [p. 174]

[Sakurai *et al.*, 2011] Sakurai, Y., Ueda, S., , Iwasaki, A., Minato, S. et Yokoo, M. (2011). A compact representation scheme of coalitional games based on multi-terminal zero-suppressed binary decision diagrams. *In 14th PRIMA*, pages 4–18. [p. 38]

[Sandholm *et al.*, 1999] Sandholm, T., Larson, K., Andersson, M., Shehory, O. et Tohmé, F. (1999). Coalition structure generation with worst case guarantees. *Artificial intelligence*, 111(1-2):209–238. [p. 32], [p. 33], [p. 34]

[Schmeidler, 1969] Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM J. Appl. Math.*, 17(6):1163–1170. [p. 23], [p. 38]

[Schrijver, 2003] Schrijver, A. (2003). *Combinatorial optimization : polyhedra and efficiency*, volume 24. Springer. [p. 33]

[Shapley, 1953] Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317. [p. 19], [p. 38]

[Shapley et Shubik, 1966] Shapley, L. S. et Shubik, M. (1966). Quasi-cores in a monetary economy with nonconvex preferences. *Econometrica*, pages 805–827. [p. 21], [p. 38]

[Shapley et Shubik, 1974] Shapley, L. S. et Shubik, M. (1974). *Game Theory in Economics*. Rand Corporation. [p. 16]

[Shehory, 1998] Shehory, O. (1998). *Architectural properties of multi-agent systems*. Carnegie Mellon University, The Robotics Institute. [p. 8]

[Shehory et Kraus, 1995] Shehory, O. et Kraus, S. (1995). Task allocation via coalition formation among autonomous agents. *In 14th IJCAI*, pages 655–661. [p. 34], [p. 38], [p. 44], [p. 48]

[Shehory et Kraus, 1996] Shehory, O. et Kraus, S. (1996). Formation of overlapping coalitions for precedence-ordered task-execution among autonomous agents. *In 2nd ICMAS*, pages 330–337. [p. 30], [p. 34], [p. 38], [p. 44], [p. 48]

[Shehory et Kraus, 1998] Shehory, O. et Kraus, S. (1998). Methods for task allocation via agent coalition formation. *Artif. Intell.*, 101(1-2):165–200. [p. 28], [p. 29], [p. 34], [p. 36], [p. 38], [p. 44], [p. 48]

[Simon, 1969] Simon, H. A. (1969). *The Sciences of the Artificial*. MIT press. [p. 7]

[Sims *et al.*, 2003] Sims, M., Goldman, C. V. et Lesser, V. (2003). Self-organization through bottom-up coalition formation. *In 2nd AAMAS*, pages 867–874. [p. 34], [p. 35], [p. 38]

[Sklab *et al.*, 2020] Sklab, Y., Aknine, S., Shehory, O. et Tari, A. (2020). Coalition formation with dynamically changing externalities. *Eng. Appl. Artif. Intell.*, 91:103577. [p. 32], [p. 38]

[Suzumura, 1976] Suzumura, K. (1976). Rational choice and revealed preference. *The Review of Economic Studies*, 43(1):149–158. [p. 16]

[Thrall et Lucas, 1963] Thrall, R. M. et Lucas, W. F. (1963). N-person games in partition function form. *Nav. Res. Logist.*, 10(1):281–298. [p. 31], [p. 38]

[Tran-Thanh *et al.*, 2010] Tran-Thanh, L., Chapman, A., De Cote, E. M., Rogers, A. et Jennings, N. R. (2010). Epsilon–first policies for budget–limited multi-armed bandits. *In 24th AAAI*, pages 1211–1216. [p. 52]

[Ueda *et al.*, 2012] Ueda, S., Hasegawa, S., Hashimoto, N., Ohta, N., Iwasaki, A. et Yokoo, M. (2012). Handling negative value rules in MC-net-based coalition structure generation. *In 11th AAMAS*, volume 12, pages 795–804. [p. 32], [p. 38]

[Vallée, 2015] Vallée, T. (2015). *De la manipulation dans les systèmes multi-agents : une étude sur les jeux hédoniques et les systèmes de réputation.* Thèse de doctorat, Université de Caen Normandie. [p. 52]

[Vallée et Bonnet, 2017] Vallée, T. et Bonnet, G. (2017). Jeux de coalitions hédoniques à concepts de solution multiples. *In 25e JFSMA*, pages 52–62. [p. 16], [p. 17]

[van Der Laan et van Den Brink, 1998] van Der Laan, G. et van Den Brink, R. (1998). Axiomatization of a class of share functions for n-person games. *Theory and Decision*, 44(2):117–148. [p. 20]

[Vercouter, 2000] Vercouter, L. (2000). *Conception et mise en oeuvre de systèmes multi-agents ouverts et distribués. (Design and implementation of open and distributed multi-agent systems).* Thèse de doctorat, École nationale supérieure des mines de Saint-Étienne, France. [p. 8], [p. 10]

[Vinyals *et al.*, 2012] Vinyals, M., Bistaffa, F., Farinelli, A. et Rogers, A. (2012). Stable coalition formation among energy consumers in the smart grid. *In 3rd ATES*, pages 35–42. [p. 38]

[Voice *et al.*, 2012] Voice, T., Polukarov, M. et Jennings, N. R. (2012). Coalition structure generation over graphs. *J. Artif. Intell. Res.*, 45:165–196. [p. 34], [p. 35], [p. 36], [p. 38]

[Wooldridge et Dunne, 2006] Wooldridge, M. J. et Dunne, P. E. (2006). On the computational complexity of coalitional resource games. *Artif. Intell.*, 170(10):835–871. [p. 28], [p. 38]

[Xu *et al.*, 2016] Xu, G., Dai, H., Hou, D. et Sun, H. (2016). A-potential function and a non-cooperative foundation for the solidarity value. *Oper. Res. Lett.*, 44(1):86–91. [p. 20], [p. 38]

[Xu *et al.*, 2017] Xu, G., Li, X., Sun, H. et Su, J. (2017). The Myerson value for cooperative games on communication structure with fuzzy coalition. *J. Intell. Fuzzy Syst.*, 33(1):27–39. [p. 28], [p. 29], [p. 38]

[Yun Yeh, 1986] Yun Yeh, D. (1986). A dynamic programming approach to the complete set partitioning problem. *BIT*, 26(4):467–474. [p. 32]

[Zick *et al.*, 2012] Zick, Y., Chalkiadakis, G. et Elkind, E. (2012). Overlapping coalition formation games : Charting the tractability frontier. *In 11th AAMAS*, pages 787–794. [p. 31], [p. 38]

Formation de coalitions répétée dans un contexte stochastique : protocoles et expérimentations

Cette thèse porte sur l'étude des jeux de coalitions stochastiques répétés qui permettent de lever certaines hypothèses fortes souvent considérées dans les jeux classiques, comme la connaissance *a priori* des utilités associées aux coalitions ou la nature détermiste de ces utilités. Dans la première partie de cette thèse, nous avons établi un modèle de formation de coalitions stochastique, et nous avons proposé un protocole d'apprentissage de la fonction caractéristique sur la base de jeux répétés. Nous avons également défini plusieurs concepts de solution fondés sur une notion d'équilibre exploration-exploitation. Des expérimentations montrent qu'un de nos concepts est aussi efficace qu'une approche gloutonne sans toutefois la surpasser. Dans la seconde partie de cette thèse, nous avons adapté un protocole connu de négociations multilatérales au cadre de la formation de coalitions classique, et proposé des stratégies mieux adaptées à ce cadre, d'abord dans un contexte distribué, puis décentralisé. Ensuite, nous avons étendu ce protocole au cas des jeux répétés et stochastiques, avec de nouvelles stratégies, pour les contextes distribué et décentralisé. Une analyse empirique a permis de montrer que nos stratégies sont efficaces dans les cadre distribués déterministe et stochastique.

Mots-clefs : Intelligence artificielle, Formation de coalitions, Théorie des jeux coopératifs, Systèmes multi-agents

Repeated coalition formation in a stochastic context : protocols and experiments

This thesis focuses on the study of repeated stochastic coalitional games which allow to lift some strong assumptions often considered in classic coalition formation, such as the *a priori* knowledge of the utilities of the coalitions or the deterministic nature of these utilities. In the first part of this thesis, we have established a model of stochastic coalition formation, and we have proposed a protocol for learning the characteristic function on the basis of repeated games. We have also defined several solution concepts based on a notion of exploration-exploitation equilibrium. Experiments show that one of our concepts is as efficient as a greedy approach however without surpassing it. In the second part of this thesis, we have adapted a well-known multilateral negotiation protocol to the framework of coalition formation, and proposed strategies that are adapted to this framework, first in a distributed and then in a decentralized context. Then, we have extended this protocol to the case of repeated and stochastic games, with new strategies, for both distributed and decentralized contexts. An empirical analysis allowed us to show that our strategies are efficient in the distributed deterministic and stochastic settings.

Keywords : Artificial intelligence, Coalition formation, Cooperative game theory, Multi-agent systems