Reaching Agreement in Multiagent Systems

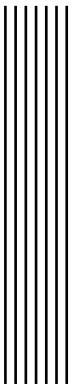
Nicolas Maudet

Reaching Agreement in Multiagent Systems

Nicolas Maudet

Coordinateur	Jérôme Lang	DR CNRS Université Paris-Dauphine
Rapporteurs	Gerhard Brewka	Professeur Université de Leipzig
	Pierre Marquis	Professeur Université d'Artois
	Jeffrey Rosenschein	Professeur Université de Jérusalem
Examinateurs	Denis Bouyssou	DR CNRS Université Paris-Dauphine
	Amal El-Fallah Seghrouchni	Professeur Université Pierre-et-Marie-Curie
	Bettina-Elizabeth Klaus	Professeur HEC Lausanne

Habilitation à Diriger les Recherches en Informatique
Soutenue publiquement le 13 Décembre 2010
Université Paris-Dauphine



Contents

Acknowledgements / Remerciements.

One day in Spring 2010, I was stuck at Porto airport (because of the ash cloud), queueing hopelessly to get the confirmation that my flight was indeed not the only one to be maintained in the whole of Europe. All of a sudden, Jérôme came appeared from nowhere, with a full alternative plan involving about thirty local trains, which would then take us through north of Spain to get to Paris in two days. I had the confirmation that this guy was born to coordinate things, and asked him right away whether he would be kind enough to also supervise my habilitation work. I think the first draft version of the document I sent him was literally like an ash cloud, and Jérôme indeed found the right tracks to reach the station in due time. Thanks for all this.

Next, I want to thank again all the members of the jury: Gerhard Brewka, Pierre Marquis, and Jeffrey Rosenschein who accepted to be reviewers ("rapporteurs") of this habilitation; as well as Denis Bouyssou, Amal El-Fallah Seghrouchni, and Bettina Klaus, for taking on their precious time to read this manuscript. It is really a great honour to have the seven of you on the front page of this document.

Over these years, I have been extremely lucky to have the opportunity to partly supervise the work of some PhD students. A good proportion of the contents of this document were either directly extracted from the works produced during these collaborations, or indirectly inspired by these very fruitful interactions. So I really want to pay tribute to all of them (by chronological order): Sylvia Estivie, Gauvain Bourgne, Wassila Ouerdane, Guillaume Ravilly-Abadie, and Dyonisios Kontarinis. Furthermore, I had the occasion to follow the work of Gael Hette, Pierre Munck, and (more remotely) Marc-André Labrie, during their Masters, and this was an equally enriching experience.

Of course I want to also thank all the colleagues I was happy to get to know and collaborate with, starting with Ulle and Yann with whom I shared most of these scientific adventures (and my offices), from the tea sessions in the common room at Imperial, to the café sessions at l'Autre Café. Then I take a big breath of air to thank my other co-authors: Leila Amgoud, Elise Bonzon, Brahim Chaib-Draa, Joris Hulstijn, Katsumi Inoue, Tony Kakas, Gabriele Kern-Isberner, Christophe Labreuche, Wenjin Lu, Jérôme Monnot, Pavlos Moraitis, Philippe Muller, Simon Parsons, Suzanne Pinson, Laurent Prevot, Iyad Rahwan, Fariba Sadri, Henry Soldano, Kostas Stathis, Francesca Toni, Alexis Tsoukias, and Lirong Xia. To this list should be appended all the colleagues involved in the different projects or research groups I belong(ed) to (MARA Meetings, FET SoCS, ANR PHAC, ANR COMSOC, COST Algorithmic Decision Theory, CREPES, EASSS, IAF, ...), too many to be listed, I hope they will excuse me for being lazy here.

Je veux également remercier Jean-Paul, César, et toutes les personnes du labo et de l'Université qui ont mis des bulles dans ces années, collègues et champagnons avec qui les problèmes de prise de décision collective les plus épineux (pause café ou pause thé?, ascenseur ou escalier?, CROUS ou brasserie?) finissent généralement en crise de dérision collective.

Je conclus en remerciant famille et amis pour la constance et la chaleur.

Résumé en français. Ce document présente de manière concise (et, je l'espère, cohérente) les recherches que j'ai menées depuis l'obtention de ma thèse dans le domaine de l'informatique, plus précisemment de l'intelligence articifielle et des systèmes multiagents.

Ces systèmes mettent en jeu des entités artificielles, conçues par des utilisateurs différents, devant se coordonner pour atteindre leur but. La problématique générale est donc l'atteinte d'états "satisfaisants" en dépit de contraintes liées à la distribution des entités qui prennent part à la décision collective, et du caractère non nécessairement coopératifs de ces agents. Le document, après une courte introduction agrémentée d'un exemple et de quelques rappels de notions utiles à la bonne compréhension du texte, se décompose en trois chapitres principaux.

- Dans le chapitre 3 ("Voting") je discute de problèmes de vote dans le cas où les profils représentant les préférences des agents prenant part à la décision ne sont pas complètement spécifiés. Cela peut être dû, par exemple, à la perte de messages du fait de la distribution, ou encore à l'impossibilité de spécifier parfaitement un profil portant sur un nombre rédhibitoire d'alternatives (comme dans le cas de domaines combinatoires). Les questions que nous abordons sont par exemple celles de la taille minimale nécessaire à encoder le profil partiel tout en restant capable de déterminer de manière certaine l'alternative choisie après complétion des votes, ou encore de la difficulté (algorithmique) liée à la détermination des alternatives que l'on peut exclure sans craindre de regretter ce choix plus tard, même si d'autres alternatives peuvent apparaîtrent.

- Dans le chapitre 4 ("Allocating Resources") j'aborde des procédures complètement décentralisées d'allocation de ressources. Le problème est complexe, en particulier du fait de synergies éventuelles entre les ressources, ce qui nécessite de représenter potentiellement les préférences sur tous les lots possibles. Ici on suppose que les agents débutent avec une allocation initiale et modifient de manière itérative cette allocation par le biais de contrats, c'est-à-dire de réallocation locale de ressources entre eux. En posant la contrainte que chacun de ces contrats doit être individuellement rationnel (au sens myope, les agents ne planifiant pas) on se penche sur les garanties de convergence de tels systèmes. En particulier, on discutera de restrictions de domaines suffisants, nécessaires ou maximaux à garantir l'atteinte d'états optimaux lorsque les échanges permis entre les agents restent simples. Les notions d'optimalité discutées incluent également des critères de justice comme l'égalitarisme ou l'absence d'envie.

- Dans le chapitre 5 ("Persuading Others") j'envisage un processus de prise de décision collective plus délibératif, au sens où les agents peuvent échanger des arguments et contre-arguments, pour éventuellement modifier le point de vue des autres. Dans un premier temps je discute d'un cadre où les agents coopèrent en vue d'établir un diagnostic commun d'une situation. Encore une fois, la difficulté vient du fait que les agents ne percoivent que localement leur environnement et ne disposent que de possibilités restreintes de communication (en particulier, tout agent ne peut pas communiquer avec tout autre agent). Dans la mesure où chaque agent construit (sur la base d'informations partielle) une hypothèse qui pourra être par la suite réfutée par d'autres agents, nous sommes en présence d'un raisonnement de type non-monotone. Enfin je présente brièvement le cadre non-coopératif d'une argumentation multi-partite, où les agents peuvent avoir des opinions réellement contradictoires. Un protocole simple est proposé, qui contraint minimalement la pertinence des arguments échangés, et quelques phénomènes liés au comportement statégique des agents sont illustrés.

1 Introduction

We are in the presence of a multiagent system when a number of autonomous artificial entities (agents) interact more or less loosely, more or less cooperatively, with the aim of achieving the objective they have been designed for [Woo09, SLB09].

Often these agents need to reach agreements [RZ94, Kra01]: among a set of possible alternatives (or candidates) they have to jointly agree upon one of these available choices. Reaching agreements is challenging because individual agents may have conflicting views (beliefs, preferences) about the issue at stake. The issue, as we shall see in this document, may be of different nature: electing a candidate, reaching an agreement about a disputed claim, allocating a number of resources among a set of agents. Further the procedure that leads to such an agreement being made can be of various types, from a purely centralized algorithm to a protocol designed for self-interested agents. Following this and in order to situate our contribution with respect to the relevant literature, we shall emphasize four distinctive features.

- *Are the agents cooperative?* The setting may indeed involve *cooperative* or *self-interested* agents. Self-interested agents are solely guided by their own interest. Typically, applications where a single designer implements the whole system (including agents) is a cooperative setting (unless it is for simulation purposes), as opposed to applications where the designer of the applications has no control over the agents' internal design.

- *What is the nature of the agreement?* An agreement may of course concern a *decision* to be taken (in this case we typically speak of a collective decision making problem), but is not limited to that: agents may want to agree on the current *state of the world*. Both aspects can occur in the same problem, an agreement on the state of the world being typically a pre-requisite for a good decision to be made at a later stage.

- *What are the characteristics of the domain?* A distinctive feature of the domain is whether the output will be the same for all agents, or whether each agent can be considered to be assigned part of the output. In the latter case, agent's preferences are typically solely based on the part assigned to them (an assumption known as the absence

of externalities). In the context of negotiation, this distinction is very similar to the one made in [RZ94] between *task-oriented* and *state-oriented* domains. Another important aspect is whether the domain is *combinatorial* (or multi-attribute), by which I mean that the alternatives defining the agreement space are defined over a combination of related features [CELM08]. In this case, the number of alternatives to consider is likely to be prohibitive for an explicit representation.

- *To what extent is the process distributed?* It is sometimes the case agents need to reach such a collective decision despite the fact that they are *distributed*. This does not necessarily mean that *no* central authority is involved in the process: different degrees of distribution can be conceived. Agents may still rely on a central authority to mediate the interaction and perform part of the computation. But they may also behave completely independently, perform local computations, and interact with others. This issue of distribution raises in turn a number of questions. The communication of agents may be limited in many respects: agents may have the ability to interact with some of the other agents only, the type of interaction may be restricted itself, and communication may be faulty. Agents may only perceive locally their environment: in particular, it may be the case that no agent can be assumed to hold a global picture of the system.

Of course these dimensions certainly do not exhaust the many variants of agreement problems that one may find in multiagent systems, let alone in more general settings (see *e.g.* [Con10] for a recent survey of collective decision problems). But they help to situate our contributions with respect to other related approaches in the literature. To start with we shall make clear that (with one exception) our work makes no assumption of cooperativeness of the agents and that it always involve some distribution of the process.

Finally, two perspectives can be taken on multiagent systems. The point of view of the (designer of) agents which seek to behave appropriately so as to achieve the goals they are designed for. Or the point of view of the (designer of) the system, whose aim is to set up a set of rules such that some properties can be guaranteed, despite the autonomy of agents. The collection of works presented here (mostly) adopts the system's designer point of view. This means that the properties proven are relevant at the level of system and concern problems that occur for the designer (*e.g.* guarantees on how the system will evolve, or on the amount of communication involved during the overall process). Before we get into the detail of these properties, we start with a simple illustrative example.

1.1 A Motivating Example

The story (inspired by the postmen domain of [RZ94]) involves a number of robots of different types, in charge of the area depicted in Figure 1.1. Robots R_1 and R_2 must take care of the targets t_1, t_2, t_3, t_4. In particular, when the targets send emergency messages, the robots must visit the targets at the earliest. However, as we shall see, the sensors of R_1 and R_2 are poorly designed, and they will have to communicate in order to determine the origin of the messages they receive.

Then there is a single robot R_3, which is in charge of the locations A, B and C. These places are populated by a number of agents (indicated in brackets on Figure 1.1). The task

of this robot is also to visit these three locations. However in this case, the tour is not chosen by R_3 itself but is specified by the agents populating the locations.

All this takes place in the same environment. Thus, an important aspect is that a segment of road (indicated by x on the picture), is potentially shared by both R_1, R_2 and R_3. This means that the decision taken by the first team of robots is not without consequence on the decision of R_3.

This made-up scenario involves several aspects that will be discussed in this document.

1. robots R_1 and R_2 have to collect and exchange information in order to decide which target they must visit;

2. robots R_1 and R_2 must agree upon a "good" allocation of targets to visit among themselves;

3. agents living in A, B and C have to vote to decide upon the tour that will be taken by robot R_3.

Let us now discuss the different stages and decision-making processes involved.

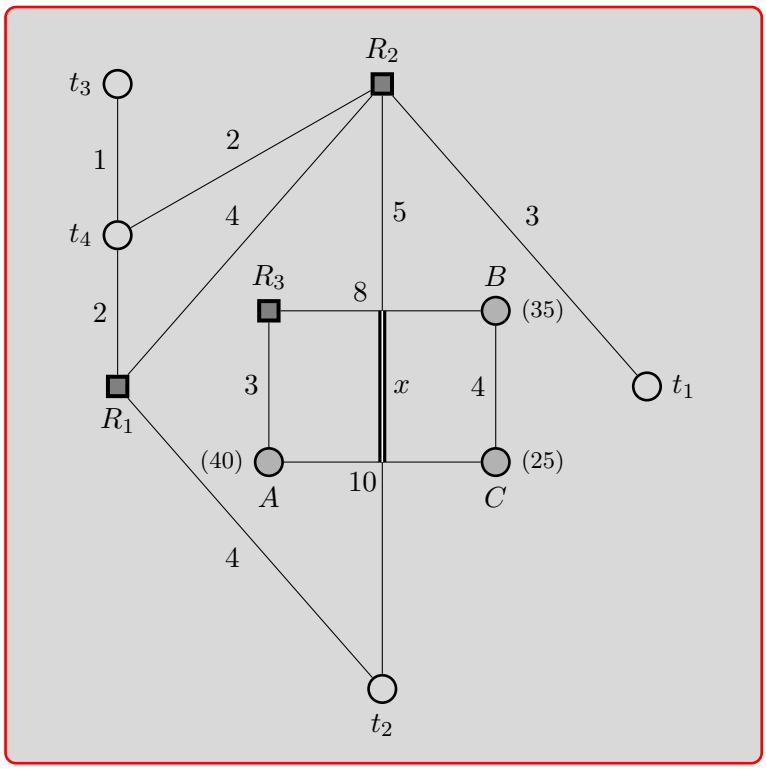


Figure 1.1: The area robots R_1, R_2, and R_3 are in charge of

(a) Robots must identify the set of targets to visit. Agents R_1 and R_2 are equipped with simple sensors: they can only perceive the direction an emergency signal comes from (when they lie on the same road as the target), but are not able to identify the distance.

As some targets are located on the same roads, robots may have to interact to identify unambiguously their objectives.

Assume R_1 receives a message coming from the "north". It can only conclude that either t_3 or t_4 sent it. Suppose that R_1 has good reasons to believe the message is more likely to come from t_4. It sends this hypothesis to R_2. This robot cannot perceive messages from t_3, but as it did not receive anything from t_4, it replies to R_1 that $\neg t_4$. The set of targets to be visited is finally identified as being $\{t_1, t_2, t_3\}$.

(b) Robots must agree upon a satisfying allocation of targets to visit. The cost for a robot when moving along an edge is indicated on Figure 1.1. Robots R_1 and R_2 have the following cost when visiting sets of targets (note that this correspond to the best possible tour for an agent, hence requiring to compute a TSP for each entry in this table).

	$\{t_1\}$	$\{t_2\}$	$\{t_3\}$	$\{t_1, t_2\}$	$\{t_2, t_3\}$	$\{t_1, t_3\}$	$\{t_1, t_2, t_3\}$
R_1	7	4	3	12	10	9	17
R_2	3	5	3	11	10	9	17

Initially, based on pre-determined areas assigned to agents, a default allocation is given: robot R_1 should visit t_3 and robot R_2 should visit t_1 and t_2. The overall cost for the team in this case is 14 (3 for R_1, and 11 for R_2: visiting t_1 then t_2). Also the last target t_2 is only visited at time 11.

Now robots may engage in a negotiation and find out that they can swap their targets t_3 and t_2. The resulting allocation is then $\{t_1, t_3\}$ for agent R_2 (for a cost of 9), while agent R_1 is assigned target t_2 (for a cost of 4). The overall cost for this allocation is 13. The last target to be visited is in this case t_1 or t_3. For both criteria considered, this allocation works better than the initial one.

(c) Agents must agree upon a route for agent R_3. Agents are truly selfish in this case: they want to be visited as early as possible (and then for routes of equal utility for them, they prefer routes which provide the shortest overall duration).

As agents assume initially that the segment of road x is busy, there are initially 4 possible options for agents, given in the following table.

o_1	ACB	17
o_2	BCA	22
o_3	ABC	18
o_4	BAC	29

However, there is some uncertainty on the availability of the segment x, whose cost is unknown for robot R_3. It depends on whether robots R_1 and R_2 will use this road to complete their own mission. If this bit of road becomes available, a new option should be considered, namely $AxBC$[1].

Given the number of agents populating each node of the area, the following profile is obtained: agents simply rank the different options available. For instance, for the 40 agents living in A, the preferred option is o_1 (the visit occurs at time 3), then o_3 (the visit occurs at the same time but the overall tour is longer), then o_4 (the visit occurs at time 19), and finally o_2 (the visit would occur at time 22).

[1]We suppose for simplicity that this road is a one-way road.

40	35	25
o_1	o_2	o_2
o_3	o_4	o_1
o_4	o_3	o_3
o_2	o_1	o_4

Suppose now that a Borda count is used to select the route of robot R_3. Under this voting rule, the score of the different options is simply computed by giving 3 points to the first option, 2 to the second, and so on. Hence we get:

$$\begin{array}{lll} o_1: & 3 \times 40 + 2 \times 25 & = 170 \\ o_2: & 3 \times 35 + 3 \times 25 & = 180 \\ o_3: & 2 \times 40 + 35 + 25 & = 140 \\ o_4: & 40 + 2 \times 35 & = 110 \end{array}$$

What does it tell us? At this point, o_2 is the preferred option, even though it is the least preferred for agents of A. But suppose that agents get to know that the segment x is now available. Does it wreck all the conclusions drawn so far? Well, even with 1 new option, o_4 is certainly not a winning option: it may already be discarded. Why is it so? Because o_1 is 60 points ahead of o_4, and the best that can happen is that the new tour o_5 would be ranked between o_4 and o_1 by agents living in B (and, say, last for agents in A and C). As there are only 35 people living there, this would not be sufficient to make o_4 the socially preferred option under this voting rule.

1.2 Properties

We now provide a more detailed description of the kind of properties that we seek to obtain when reaching agreements in multiagent systems. Again, a clear source of inspiration is [RZ94]. First, it is important to make a clean distinction between two conceptually different notions: *protocols* (of the system) and *strategies* (of agents). Protocols are supposed to be publicly known, while strategies are private to the agents. At this point it is enough to assume that agents will have preferences over the states of the system (a reflexive, transitive and unless stated otherwise complete binary relation over these states, or agreements), and that they will act in a way that (they think) will allow them to reach states that as good as possible to them in the context of this interaction (they are *individually rational*). This raises an immediate question that we discuss in priority: what is the range of possible actions that agents can take? In particular, are they restricted to what is allowed by the protocol? Are they restricted in the sense, for instance, that they would not lie to other agents? Once this is done we make more precise the notions that can be used to assess the *properties of states* of the system, from the global (designer's) point of view. In particular we informally introduce notions of *efficiency* and *fairness*. Finally, at a higher level, we introduce the typical properties worth investigating in such systems, regarding in particular the dynamics and the communication complexity involved.

1.2.1 The Conformance Problem

In isolation, a protocol cannot really provably guarantee much... What can be shown is that it does not prevent agents from reaching certain states, that it does not allow loops, etc. As

to which states are actually reached, well, it depends on the way agents will behave. For instance, as an extreme case, we may not prevent agents from remaining silent, whichever moves are allowed by the protocol. Throughout this document, we will then assume that an agent always sends one and only one answer among those allowed by the protocol to any given message, except when it is in a final state of the protocol (in which case it does not have to send an answer). This property corresponds for instance to both *determinism* and *exhaustiveness* in [STT01].

Doing so, we assume that agents' strategies conform to the protocol. At this point we should emphasize that this constitutes a strong assumption. In distributed systems, the problem of actually verifying that this property indeed holds, either on-the-fly (when moves are played) or a-priori (by inspecting agents' strategies against a given protocol) is non-trivial. A key to this problem is whether the representation of protocols allows agents to reason about these rules[2].

In [EMST03b, EMST04] (joint work with Ulle Endriss, Francesca Toni, and Fariba Sadri) we investigate the question for a certain type of logic-based representations of protocols (consisting of translating protocols as integrity constraints in abductive logic programs). In collaboration with Tony Kakas and Pavlos Moraitis [KMM05], we touch the same issue but build on an alternative method to represent protocols and strategies in the argumentation-based framework of [KMD94, KM03] (which offers a great modularity at design time, and a great flexibility at run-time: for instance, one can easily specify policies that are adaptable depending on specific circumstances that may occur, as specified by the designer). We will not elaborate further on these issues here, and simply assume that agents follow the rules of the game. We just note that one reason why the semantics of agent communication languages should avoid referring to mental states is precisely because verifying conformance in this case is difficult [Woo00, MCd02].

1.2.2 The Manipulation Problem

Many protocols will contain moves that require agents to reveal (part of) their preferences, or beliefs: for instance, we may ask an agent to tell how much they would be prepared to pay to obtain a given item, or to cast a ballot when they vote, or to furnish a counter-argument to a claim if they can, etc. Intuitively, a move is then *sincere* when it is in accordance with the agent's internal state. Sometimes however the agent may realize that he would be better off not playing sincere, in the sense that he would prefer the agreement reached by misreporting, typically, his preferences (but the question is also meaningful for beliefs [EKM07]). In this case we are in presence of a *manipulation*. In game-theory, the approach of *mechanism design* has investigated this question in detail, developing in particular protocols that have the property to be *truthful*, that is, protocols for which there is an incentive for agents to play truthfully (in the sense that this is a dominant strategy). One may wonder why this property is sought from the designer's point of view, especially when one considers that the approaches that ensure truthfulness usually comes at a price since they rely on designing payments to agents

[2]Pushing this idea a bit further leads to the perspective of a complete argumentation-based agent architecture, where the different modules composing an agent's behaviour are implemented as argumentation-based decision processes, and more importantly where the interplay and arbitration between these different modules is also a matter of an argumentation-based decision process. This (ambitious) agenda was set up during a Dagstuhl seminar on argumentation theory, and together with Leila Amgoud, Tony Kakas, Gabriele Kern-Isberner, and Pavlos Moraitis, we recently did put forward a first proposal that looks in that direction [KKIA+10]

that make manipulation non profitable. One reason is that the cost of agents strategizing can be very high and indeed lead to very inefficient states.

1.2.3 Efficient States

A first rather weak definition of the social efficiency of a state is to say, following the Pareto criteria, that no other state should be equally preferred by all agents and strictly preferred by at least one of them. This criterion is rather uncontroversial, but it is not very selective: for instance, if two agents want to share a cake (and they would like the biggest possible slice), then any way to share the cake in two pieces results in a Pareto-efficient state. A related problem is that there may be a very large number of Pareto-efficient states, so it is usually required to apply a more restrictive criteria. A nice property of this criterion though is that it does not require the interpersonal comparison of preferences among agents: it thus remains a very useful criterion in domains where comparing the intensity of preferences between agents cannot be justified. When such comparison is possible, we may instead seek to maximize the average satisfaction in the society following the well-known *utilitarian* principle, or the product of these intensities as advocated by the Nash product.

1.2.4 Fair States

Fairness, on the other hand, evaluates how well "balanced" is a state, whether some agent's preference is obviously disregarded. One may for instance pay special attention to the least happy agent in the society. More generally, one may use various *inequality indices* to evaluate how "fair" is the vector reporting the respective satisfaction of the different agents. In the context of voting, this may be interpreted as discarding candidates that are ranked very low by some voters, and give preference to more consensual candidates. In the context of agents with conflicting beliefs, this motivates [KP05] the use of appropriate distances when merging their bases. When the output can be differentiated among the agents, another approach is instead to rely on the subjective evaluation, by agents themselves, of the fairness of the state: this leads to notion of fairness based on how *envious* are agents of the share obtained by other agents. Interestingly, these notions of fairness can sometimes be "hidden" and be given a different interpretation: consider the case of an agent buying goods from a set of providers. Perhaps there is some uncertainty on the reliability of these providers, in which case the agent may want to secure a *robust* deal by diversifying the providers to contract with.

1.2.5 Properties of Dynamic Systems

Finally there are properties of the system itself, taken as a whole and considering its dynamics (when applicable). Here we assume that the system evolves from state to state as a consequence of the moves of the agents. Then there are obvious properties that one may be interested to investigate: (*i*) *termination:* does the system reach a state where no further move is possible?; (*ii*) *convergence:* does *any* sequence of moves lead to some state with a given property?; (*iii*) *reachability:* does *some* sequence of moves lead to some state with a given property?

Note that convergence is typically a property that the designer wants to implement, while reachability is arguably more relevant for individual agents. When no further criteria is given to indicate when the system should stop (as, for instance, a limited number of rounds),

the system stops when no agent has any further rational move to play, and termination corresponds in that case to *stability*. A popular notion, introduced in recent years by [KP99] is the *price of anarchy* (resp. *stability*): it gives a quantification of the ratio of the worst (resp. best) stable state over the socially optimal state. In other words, it gives a measure of what the designer loses by letting agents act as self-interested entities.

1.2.6 Communication Complexity

Unfortunately, distribution is no magic: what is (sometimes) gained in reducing the burden of the computational tasks that rest on each individual agent, often comes at the price of the communication overhead that may result for the overall system. It is then of the utmost importance to evaluate the communication requirements of the different system designed. Depending on the problem at hand, we may be interested in studying, for example, the complexity of the language required to be able to meaningfully interact, the rate of convergence in case of a dynamic system, and ultimately the number of bits exchanged in all cases. It is then possible to analyse the communication requirement of a given protocol, either in the worst-case or on average (typically by means of experiments in the latter case). But it is also possible to investigate the *communication complexity of a problem*, which provides a more fundamental understanding of the communication burden induced: specifically, it allows to (lower) bound the amount of communication required by the best conceivable protocols (in the worst-case).

1.3 Structure and Content of the Document

Now that we have gained a better understanding of the properties that shall be discussed in this document, we are in position to give the detail of the specific contexts discussed in the next chapters.

- In chapter 2 ("Voting") we take up the classical research agenda of voting. However, stimulated by the prospect of distributed systems as discussed previously, we focus our attention on situations where profiles (representing agents' preferences) may only be partially specified. This may be due to delays or loss of messages when conveyed to the central authority in charge of the computation of the outcome of the vote. A typical case is illustrated by step (c) in our example: here the set of options is not fully specified from the beginning. Alternatively, the votes of some agents may not be known initially. Several interesting problems occurs: for instance, what is the size of the minimal message that encodes a profile where voters are missing? Can we design efficient algorithms that solve potentially difficult decision problems like the "can we certainly thrown this option away" that faces the agents in our example? These issues fall under the umbrella of the emerging "computational social choice" field (see [CELM07] for an outdated survey).

- In chapter 3 ("Allocating Resources") we turn our attention to resource allocation problems, and study what can be achieved by protocols that do not rely on any central authority. In this setting, agents modify the current allocation of resources in the system by implementing local reallocation of resources between them (*i.e.* contracts). An example is provided by step (b) in our example, when robots swap the tasks that were

initially allocated to them. Sequences of such contracts constitute paths between allocations. Different important questions that a designer would like to solve arise in this context. For instance, will all sequences converge to an optimal allocation (convergence), whatever the initial allocation considered?

- In chapter 4 ("Persuading others"), we move on to a more deliberative view on agreement seeking. We now allow protocols that include the explicit exchange of arguments and counter-arguments. Of course this only makes sense if agents are prepared to modify their beliefs and preferences accordingly, on the basis of the information received from other agents. These types of protocols are exploited in two different situations: we first discuss the case of agents holding incomplete knowledge of their environment, leading them to make hypothesis that only hold defeasibly and can be contradicted by counter-examples communicated by other agents. A very simple illustration of this is given by step (a) in our example. We then give some insights regarding the challenging issue of designing a proper multiparty persuasion protocol, when several agents with potentially truly contradictory opinions are to come to an agreement.

The document is based on a collection of papers, the vast majority of which are already published and available. Many of these works have also been conducted in the context of some PhD (co-)supervision. Specifically, issues of voting with incomplete profiles have been studied in the context of the PhD thesis of Guillaume Ravilly-Abadie thesis (co-supervision with Yann Chevaleyre and Jérôme Lang). The fairness issues of the distributed resource allocation framework presented in Chapter 3 have mainly been studied as part of the PhD of Sylvia Estivie (co-supervision with Yann Chevaleyre). The distributed hypothetical reasoning problem discussed in Chapter 4 was one of the main question studied in the PhD of Gauvain Bourgne (co-supervision with Suzanne Pinson). Dyonisios Kontarinis is now starting a PhD (co-supervision with Elise Bonzon) on the questions of multiparty argumentation presented later on in the same chapter. Finally, the work of Wassila Ouerdane (co-supervision with Alexis Tsoukias) takes a perspective (decision-aiding processes) which is slightly off the tracks of the main story told here. I will however mention possible connections to some of the issues raised in the context of this work.

My main effort has been to put the material in perspective, to draw new connections between (part of) these works, to add examples, or to shed a new light on some results. Sometimes a sketch of proof is provided when I thought it was useful for a proper understanding of a key notion, but in general it contains very few technical details. The perspectives open by some of these works are discussed in the relevant chapters. Finally, this document is certainly not intended to be a survey of the field.

2 Voting

Properties of systems where agents (but maybe not all of them) vote on candidates (but maybe not all of them).

Voting is a primary means to reach agreement among a group of agents. In short, a voting rule maps a collection of preferences of agents over several candidates (a profile) to a single, winning, candidate. The mathematical study of voting systems goes back to [dC85], and the modern school of social choice is usually associated to the seminal work of Arrow [Arr51]. This line of work is largely axiomatic: characterizing voting procedures in terms of properties they meet, showing the impossibility for any rule to satisfy simultaneously a given set of desirable properties (the typical result in this vein being the famous theorem by Arrow himself). In the late 80's, the works of Bartholdi *et al.* [BTT89b, BTT89a] and Hudry [Hud89] were the first to import computational issues into the field of social choice, as it appeared that some questions required non-trivial algorithms to be solved. The agenda has since (in fact, very recently) been taken up by several researchers, leading to the emergence of a *computational social choice* sub-field, crossing the boundaries of AI, social choice, and operation research. The typical questions that arise are related to the following (non mutually-exclusive) aspects: (*i*) the computation of voting rules, (*ii*) the computational aspects of the strategic behaviour of agents, (*iii*) the choice of appropriate representation languages, in particular for agents' preferences defined over a very large number of options [Lan04]; or (*iv*) the design of communication protocols[1]. In this chapter we shall encounter in particular two prominent complexity issues that arise in decision problems occurring in social choice, specifically when it comes to voting:

- *Computational complexity.* This includes in particular determining the winner of an election (there are rules for which this is hard), manipulating the election (misreporting the preferences so as to get a better outcome) by an individual or a coalition. In the first case complexity is to be avoided, but in the second it may be considered as a barrier against strategic behaviour, assuming agents have limited computational abilities.

[1] We shall clearly state here that computational social choice is certainly not limited to these aspects, and refer to our survey paper [CELM07] for a more comprehensive, albeit not exhaustive, survey.

- *Communication complexity.* The question is here to determine the minimal amount of information that needs to be exchanged between agents to solve a given problem, regardless of the computational burden that rests on them. Conitzer and Sandholm [CS05] were the first to attack this problem, for the case of determining the outcome of the election (this provides insights for the elicitation protocols that can be designed, for example).

A further motivation for computer scientists to look into these questions was provided by the fact that technological advances suggested that voting could occur in situations different from the voting setting typically studied in social choice. Recently indeed, voting has also been advocated as a suitable means to reach agreement among artificial agents, in many different contexts. The following examples are rightly considered (see for instance [FP10]) as significant examples of the use of voting methods in systems as diverse as groups of coordinating agents [ER91], aggregation of search results over the web [DKNS01], or collaborative filtering for personalized recommendations [PHG00]. On top of that, with the development of online voting systems such as Doodle® (typically used for non-critical decisions), one sees processes that spans over a few days, and where voters may be informed of previously casted ballots. This also stimulates original research on voting where usual assumptions are relaxed (see for instance [MPRJ10] for a study of the convergence of voting systems where voters can repeatedly change their vote when observing the current outcome).

In many of these contexts, it is certainly natural to consider that preference profiles can be incomplete, because communication may be faulty or too restricted to allow a complete specification, as is typically the case in combinatorial domains[2].

2.1 Background: Voting under Incomplete Knowledge

2.1.1 Voting Rules

The canonical setting of voting theory is as follows. Let C be a finite set of *candidates* (or *alternatives*) and $\mathcal{N}$ be a finite set of *voters*, with $p = |C|$ and $n = |\mathcal{N}|$. A *vote* is a linear order over C. We usually denote votes in the following way: $a \succ b \succ c$ (or abc for short). A *profile* is a collection $P = \langle V_1, \ldots, V_n \rangle$ of votes. Let $\mathcal{P}$ be the set of all votes and therefore $\mathcal{P}_C^n$ be the set of all n-voter C-profiles.

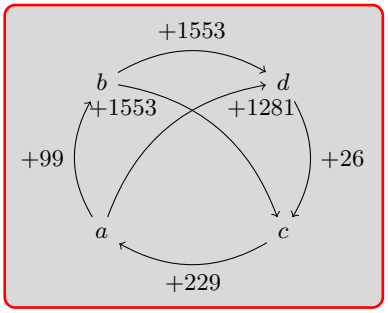
A voting rule on C is a function r from $\mathcal{P}_C^*$ to C. It thus returns a single winner. (When several co-winners are allowed we talk of voting *correspondences*). As the usual definition of most voting rules does not exclude the possibility of ties, we assume these ties are broken by a fixed priority order on candidates.

Let $\vec{s} = \langle s_1, \ldots, s_p \rangle$ be a vector of integers such that $s_1 \geq \ldots \geq s_p$ and $s_1 > s_p$. The scoring rule $r_{\vec{s}}(P)$ induced by $\vec{s}$ elects the candidate maximizing $score_{\vec{s}}(x, P) = \sum_{i=1}^{p} s_i \cdot n(P, i, x)$, where $n(P, i, x)$ stands for the number of times that x is ranked in position i in the profile P. The *plurality* rule r_{Pl} is the scoring rule corresponding to the vector $\langle 1, 0, \ldots, 0 \rangle$. The *Borda* rule r_B is the scoring rule corresponding to the vector $\langle p-1, p-2, \ldots, 0 \rangle$. The *veto* rule r_V is the scoring rule corresponding to the vector $\langle 1, \ldots, 1, 0 \rangle$. If K is a fixed integer then *K-approval*, r_K, is the scoring rule corresponding to the vector $\langle 1, \ldots, 1, 0, \ldots, 0 \rangle$—with K 1's and $p - K$ 0's.

[2]Incompleteness is certainly not limited to these contexts: for instance, incompleteness on the votes can be *intrinsic*, when it does not make sense to compare some options, see [CELM07].

A *Condorcet winner* is a candidate preferred to any other candidate by a strict majority of voters. As there does not necessarily exist one (because cycles may occur), several rules have been proposed, based on the same idea of pairwise comparison of candidates.

Let $N_P(x, y)$ be the number of voters in the profile P preferring x to y. The *majority graph* M_P is the directed graph whose set of vertices is X and containing an edge from x to y iff a strict majority of votes in P prefers x to y, i.e., if $N_P(x, y) > N_P(y, x)$. The *weighted majority graph* $\mathcal{M}_P$ is the same as M_P, where each edge from x to y is weighted by $N(x, y)$ (note that there is no edge in $\mathcal{M}_P$ between x and y if and only if $N_P(x, y) = N_P(y, x)$.) A voting rule r is *based on the majority graph* (abridged into "MG-rule") if for any profile P, $r(P)$ can be computed from M_P, and *based on the weighted majority graph* (abridged into "WMG-rule") if for any profile P, $r(P)$ can be computed from $\mathcal{M}_P$.



1340: $a \succ b \succ c \succ d$
827: $b \succ a \succ d \succ c$
75: $d \succ a \succ b \succ c$
961: $c \succ a \succ b \succ d$
1400: $d \succ b \succ c \succ a$
100: $d \succ c \succ a \succ b$

Figure 2.1: A profile and the corresponding weighted majority graph

A *Condorcet-consistent rule* is a voting rule electing the Condorcet winner whenever there is one. An important result in social choice, due to Fishburn [Fis73], shows that any positional scoring rule fails to be Condorcet-consistent. Two famous examples of rules based on the (W)MG, and Cordorcet-consistent, are:

- the *Copeland* rule, where a candidate x receives 1 point each time it beats a candidate y in the pairwise election (that is, when $N_P(x, y) > N_P(y, n)$), and half a point when it ties. The elected candidate is the one maximizing the sum.
- *Simpson* (aka *Maximin*) rule, where we pick the candidate beaten by the smallest margin in any pairwise comparison (that is, we select the candidate x that minimizes $max\{N_P(y, x)|y \neq x\}$).

Then there are rules that are performed in successive rounds. Here again two prominent examples are:

- the *plurality with runoff* rule, which consists of two rounds: the first round keeps the two candidates with maximum plurality scores (with some tie-breaking mechanism), and the second round is the majority rule between the remaining candidates.
- the *single transferable vote* (STV) rule performs in $|C|$ successive rounds: at each round, the candidate with the lowest plurality score gets eliminated and its votes are transfered to the next preferred candidate in each ballot.

This is far from exhausting the vast zoo of voting rules, but is sufficient to illustrate the results presented in this document. We finally note that Xia and Conitzer recently introduced the

notion of generalized scoring rules [XC08b]. This class corresponds to rules that assign, for each vote, a vector of score (the vector may have a number of components different from p). Then the candidate is selected by summing up these vectors and comparing the resulting scores. In fact, it has been recently shown [XC09] that this class of rules is characterized by a property called *finite local consistency* that among common rules, only Dodgson's rule is known not to meet. This idea has close connection to the notion of compilation complexity as we will see later.

2.1.2 Voting with Incomplete Profiles

Now let us consider a profile only partially specified: suppose there are three candidates a, b, and c, but we only know that $a \succ b$ and $a \succ c$. There are two possible completions of this partial vote, either $a \succ b \succ c$ or $a \succ c \succ b$. Perhaps other votes are also partially specified wrt. other candidates.

In such a situation, a natural question arises: given a partial profile, is it the case that a candidate is *possible winner*. These notions have been introduced by Konczak and Lang in [KL05]: given a collection $\langle V_1, \ldots, V_n \rangle$ of partial strict orders on C representing some incomplete information about the votes, a candidate x is a possible winner if there is a profile $\langle V'_1, \ldots, V'_n \rangle$ where each V'_i is a ranking on C extending P_i, in which x wins. Similarly, the notion of *necessary winner* can be defined when all such profile extensions make x winner.

This question, in this general form, has motivated a great amount of works: the computational complexity of the problem consisting in deciding whether a candidate is a possible (resp. necessary) winner has been studied in detail in [XC08a, PRVW07]. The parametrized complexity of these problems has also been studied [BHN09], and the counting variant has also been recently considered [BBF10].

Our contribution in this rapidly expanding sub-topic of computational social choice concerns two special cases of this general problem, namely (*i*) when incomplete profiles occur because only a subset of voters have casted their ballots already (joint work with Yann Chevaleyre, Jérôme Lang, and Guillaume Ravilly-Abadie), and when (*ii*) incomplete profiles occurs because only a subset of candidates have declared themselves (joint work with Yann Chevaleyre, Jérôme Lang, Jérôme Monnot, and recently Lirong Xia). The questions that we address are of different nature.

2.2 Missing Voters: Compilation Complexity of Incomplete Profiles

The special case of incomplete profiles with missing voters is especially important. First, note that the possible winner problem is in this case equivalent to the problem of deciding whether a coalition of agents can coordinate to ensure that a given candidate is elected. The computational complexity of this problem has been studied widely under the name *coalitional manipulation* [CLS03], and several algorithms (exact or approximations) have been proposed in the literature (see [FP10] for a recent survey on the complexity of manipulation issues). A very related problem is the *complexity of vote elicitation* [CS02, Wal08]: given a voting rule r, a set of known votes P^m, and a set of t new voters, is the outcome of the vote already determined from P^m (in other words, is there a necessary winner for the election at this stage of the elicitation process)?

In [CLMRA09] we indeed address a different, but complementary problem: what is the minimal amount of space needed to synthesize the information contained in the votes of the subelectorate while keeping enough information for computing the outcome once the last votes are known. We dubbed this notion the compilation complexity of voting rules[3]. Interestingly, this problem can also be interpreted as determining the *one-round* communication complexity of these voting rules: suppose indeed we have two agents, one dedicated to collect all the m votes known so far, while the other one is dedicated to collect the remaining vote (and compute the outcome). We are after the minimal size of the message to be send from A to B.

Definition 2.2.1 *Given a voting rule r, we say that a function σ from $\mathcal{P}_X^m$ to $\{0,1\}^*$ is a* compilation function *for r if there exists a function $\rho : \{0,1\}^* \times \mathcal{P}_X^* \to X$, such that for every $P \in \mathcal{P}_X^m$, every $k \geq 0$ and every $R \in \mathcal{P}_X^k$, $\rho(\sigma(P), R) = r(P \cup R)$. The* compilation complexity *of r is defined by*

$$C(r) = \min\{Size(\sigma) \mid \sigma \text{ is a compilation function for } r\}$$

Let us start with some general and intuitive remarks. A universal upper bound is certainly $m \log(p!)$, by binary encoding the integer value assigned to each possible vote of each voter. But for anonymous rules, one may instead, for each possible vote, encode the number of voters that chose it.

Proposition 2.2.1 *Let r be an anonymous voting rule, m be the number of initial voters and p the number of candidates. Then $C(r) \leq min(m \log(p!), p! \log m)$.*

Let us consider the example profile given in Figure 2.1.1. There are 4703 voters and four candidates, hence 24 possible votes. Here we get $min(4703 \log 24, 24 \log 4703)$: the anonymous compilation is clearly more efficient (it requires 312 bits *vs.* 23515 bits for the non-anonymous compilation). Interestingly, this anonymous encoding of a profile is known in social choice as a *voting situation* [BL92].

Intuitively, the message can be shorter than naively encoding the whole profile, even anonymously, because the voting rule may disregard part of the input and consider some profiles as equivalent. We make this idea more precise.

Let $P, Q \in \mathcal{P}_X^m$ be two m-voters X-profiles and r a voting rule. P and Q are *r-equivalent* if for every $k \geq 0$ and for every $R \in \mathcal{P}_X^k$ we have $r(P \cup R) = r(Q \cup R)$.

Example 1 *Let r_P be plurality and r_B Borda, $X = \{a, b, c\}$ and $m = 4$. Let $P_1 = \langle abc, abc, abc, abc \rangle$, $P_2 = \langle abc, abc, acb, acb \rangle$, and $P_3 = \langle acb, acb, abc, abc \rangle$.*

- *P_2 and P_3 are r_{Pl}-equivalent and r_B-equivalent. More generally, they are r-equivalent for every anonymous voting rule r.*
- *P_1 and P_2 are r_{Pl}-equivalent. However they are not r_B-equivalent: take $R = \langle bca, bca, bca \rangle$, then we have $r_B(P_1 \cup R) = b$ while $r_B(P_2 \cup R) = a$.*

We can now provide the following characterization of $C(r)$. Up to minor details, this is a reformulation (in our own terms) of Exercise 4.18 in [KN97].

[3]Following the ideas developed in the area of knowledge compilation [CDLS02, DM02], the objective could also be to compile the information, using as much off-line time and space as needed, in such a way that once the last votes are known, the outcome can be computed as fast as possible.

Proposition 2.2.2 *Let r be a voting rule, m be the number of initial voters and p the number of candidates. If the number of equivalence classes for $\sim_r$ is $g(m,p)$ then the compilation complexity of $C(r) = \lceil \log g(m,p) \rceil$.*

The key is then to provide for a given rule a characterization of the equivalence class. Once this is done, the problem then boils down the enumerate the number of equivalence classes that can be found for each voting rule: sometimes this can be done easily and we provide the exact compilation complexity, often $g(m,p)$ is difficult to get exactly and we seek upper and lower bounds that match asymptotically. Upper bounds are simple to obtain by simply evaluating the space required to store the information required by the characterization, for instance partial scores for positional scoring rules. Take Borda: with n agents, the Borda score of a candidate is certainly at most $n(p-1)$, and as the score of each candidate must be stored we are sure that $C(r) \leq (p-1)\log n(p-1)$. Why is this not tight? Because many of the score profiles counted here do not correspond to any *feasible* profile: for instance as soon as one candidate gets 0 we know for sure that no other score can be lower than m. The tricky bit is here (since enumerating the number of feasible profiles is hard in most cases): exhibit a sufficiently significant set of feasible profiles so as to obtain a matching lower bound.

The results presented in [CLMRA09] apply this methodology to positional scoring rules such as plurality and Borda, but also to the wide family of rules based on the pairwise comparison of candidates. In that case two profiles belong to the same equivalent class when their majority graphs correspond (this characterization is exact as long as we require rules to be Condorcet-consistent). Finally we treat the case of the *single transferable vote*, for which it is perhaps less easy to get an intuition as to what information should be stored, and show that it is necessary and sufficient to store the number of times each candidate would be ranked first in the absence of any possible subset of other candidates. Interestingly, these results suggest a ranking among voting rules which significantly differs from the results regarding the communication complexity [CS05]. Table 2.2 summarizes the characterization of equivalence classes and the obtained compilation complexity results for these rules[4].

Voting rule	*Characterization of equiv. classes*	*Compilation complexity*
Any voting rule	same profiles	$\Theta(mp\log p)$
Anonymous	same voting situations	$\Theta(p!\log m)$
STV	for all $Z \subseteq C$ and $x \notin Z$, $score_{Pl}(x, P^{-Z}) = score_{Pl}(x, Q^{-Z})$	$\Omega(2^p \log m)$ $O(p2^p \log m)$
Plurality w. runoff	$\mathcal{M}_P = \mathcal{M}_Q$ and $score_{Pl}(x,P) = score_{Pl}(x,Q)$	$\Theta\left(p^2 \log m\right)$
Cond. WMG	$\mathcal{M}_P = \mathcal{M}_Q$	$O\left(p^2 \log m\right)$ $\Omega\left(p^2 \log(\lfloor m/qp \rfloor - 2\right)$
Borda	$score_B(x,P) = score_B(x,Q)$	$\Theta(p\log m)$
Plurality	$score_{Pl}(x,P) = score_{Pl}(x,Q)$	$\Theta\left(p\log(1+\frac{m}{p}) + m\log(1+\frac{p}{m})\right)$

Table 2.1: Compilation complexity of voting rules

This set of results has been recently significantly expanded by the work of Xia and Conitzer [XC10a]. More precisely, they look at other types of compilation complexity and other rules: while our results suppose that the number of voters yet to come is unknown, they investigate

[4]It includes a correction on the lower bound of WMG rules, as provided by [XC10b]. Note also that P^{-Z} stands for the profile where the candidates Z have been deleted.

the case where the compilation complexity depends on this number, as well as the case where it depends both on the size of both sub-electorates. Some interesting questions remain open though: for instance, one may wonder whether some common voting rule need to store the full anonymous profile.

Of course in domains where the number of candidates is low, it is not very costly anyway to store all the possible orderings, and the universal compilation complexity mentioned above can be acceptable. This may not be always the case, and it is interesting to note for instance that Xia and Conitzer have used these compilation techniques in [XC10b] to optimize the computation of backward-induction outcomes in the context of Stackelberg games (and that they witnessed a significant speed up of computation times). More generally, a perspective of research could be to use a similar technique in other problems requiring to search over a set of states that can be compiled using the techniques described here.

2.3 Missing Voters: Experiments on Possible Winners

We now briefly report on a series of experiments performed as part of the PhD of Guillaume Ravilly-Abadie in order to appreciate how the average number of possible winner(s) evolves wrt. the completion of the profile of voters. These experiments involve a lot of computation, since for each partial profile one must be capable to compute which candidates are possible winners. For each candidate to be tested, the question is equivalent to that of asking whether the missing voters can form a coalition to ensure this designated candidate wins. As mentioned already, for many rules this coalitional manipulation problem is hard, although approximation can still be provided [ZPR09]. In these experiments we rely heavily on integer linear programming (ILP) formulations of the problem and used off-the-shelf solvers, which proved in most cases efficient enough to provide the desired results. Also relevant to this question is the observation made by Procaccia and Rosenschein [PR07] as well as [XC08b], namely that coalitions of size $\Theta(\sqrt{n})$ play a key role in this analysis. Indeed under this size the probability of manipulation is very low. Contrariwise, for larger coalitions manipulation seem to exist with very high probability.

Any experimental study in social choice faces the difficult problem of choosing a specific method to generate profile instances (or *cultures* the votes are drawn from). In these experiments, we make use of four different cultures: the *impartial* culture (IC) for which all possible profiles are equally probable, the *impartial and anonymous* culture (IAC) for which the voting situations are equally probable, and two types of *single-peaked* cultures. In the first one, we simply generate votes by a uniform distribution over votes that respect the single-peaked constraint (for a given order) (SP). In the second one, the peak is first chosen by a uniform distribution (PSP).

One of the main objective was precisely to compare these different cultures as far as the average number of possible winner was concerned. Figure 2.2 shows an example of the results obtained, for the plurality rule with 4 candidates, given here to illustrate the kind of observations that can come out from these experiments. First observe that for plurality, as only the top ranked candidate is relevant, the results for PSP and IC coincide. Until 80 voters, we see that no candidate among 4 (see 2.2) is eliminated under these cultures. For comparison, with 83 ballots in, while IC and PSP still do not discard any candidate, there is an average of 3 possible winners under IAC and nearly 2 under SP. The shape of SP is explained as follows: under this culture, a voter is likely to put the middle candidate at

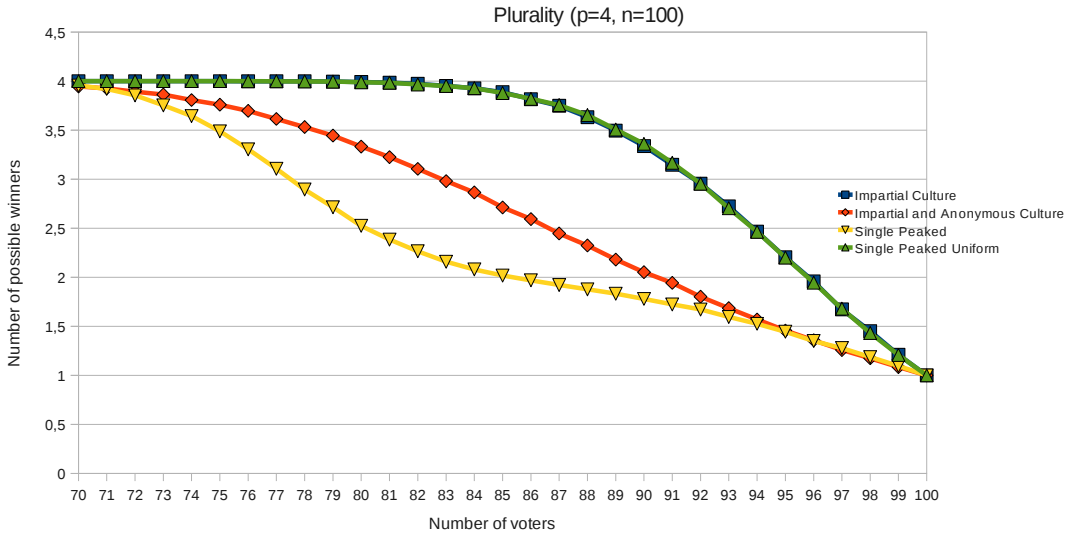


Figure 2.2: Expected number of possible winners (Plurality)

the top, and indeed we observe that two candidates (the extremes) are quickly eliminated. We expect to obtain soon a substantial catalogue of results of the like, allowing a precise comparison of voting rules and cultures involved.

2.4 Missing Candidates: the Possible Winner Problem

We now move on to a different special case of the possible winner problem with partial profiles. Here we shall consider that only a subset of candidates are initially known. Voters cast their ballots on these candidates only. We will be interested in determining who among the initial candidates are possible winners, given that a (fixed) number of new candidates may show up. The identity of these new candidates is not known yet, so no assumption can be made as to where they will be ranked when they show up. One can see that this is also a special case of voting with incomplete profiles, where the incompleteness only affects the set of candidates. If P is a C-profile and $C' \subseteq C$, then the projection of P on C', denoted by $P^{\downarrow C'}$, is obtained by deleting all candidates in $C \setminus C'$ in each of the votes of P, and leaving unchanged the ranking on the candidates of C'. For instance, let us take $P = \langle abcd, dcab\rangle$ then $P^{\downarrow\{a,b\}} = \langle ab, ab\rangle$ and $P^{\downarrow\{a,b,c\}} = \langle abc, cab\rangle$. From now on the set of initial candidates is denoted by $X = \{x_1, \ldots, x_p\}$, the set of the k new candidates is denoted by $Y = \{y_1, \ldots, y_k\}$. If P_X is an X-profile and P an $X \cup Y$-profile, then we say that P extends P_X if the projection of P on the candidates on X is exactly P_X. For instance, let $X = \{a, b, c\}$, $Y = \{d\}$; the profile given in Figure 2.1.1, repeated below on the righthandside, is an extension of the X-profile P on the lefhandside (where 1340 among the voters who voted the order $a \succ b \succ c$ chose to insert d in last position, while 75 chose to insert him in first position, etc.)

We are after (hopefully) efficient algorithms that solve this problem.

Definition 2.4.1 *Given a voting scenario $\Sigma = \langle N, X, P_X, k\rangle$, a set of new candidates Y (with $|Y| = k$), and a collection r of voting rules, we say that $x^* \in X$ is a* possible co-winner with respect to Σ, Y, and r *if there is an $X \cup Y$-profile P extending P_X such that $x^* \in r(P)$.*

$$\begin{array}{lll} 1415: & a \succ b \succ c \\ 827: & b \succ a \succ c \\ 1061: & c \succ a \succ b \\ 1400: & b \succ c \succ a \end{array} \Rightarrow \begin{array}{lll} 1340: & a \succ b \succ c \succ d \\ 827: & b \succ a \succ d \succ c \\ 75: & d \succ a \succ b \succ c \\ 961: & c \succ a \succ b \succ d \\ 1400: & d \succ b \succ c \succ a \\ 100: & d \succ c \succ a \succ b \end{array}$$

At this point it is important to note that, as a special case of the general "possible winner with incomplete profile" problem, one the one hand we inherit any easiness result from the general case (although there may be more efficient algorithms for our problem), and on the other hand any hardness result in our case strengthens hardness results of the general case. And of course the hope is to identify rules for which there is a complexity reduction for the case we consider. This problem is also highly related to manipulation by "candidate cloning". The main difference is that cloning requires a candidate and its clones to be contiguous in the votes. This problem is considered in Elkind et al. [EFS10]. Finally, there exists a special case of manipulation studied in the literature [BTT92, FHH09], where the *chair of an election* tries to control the result by adding new candidates, which is reminiscent of our problem. The difference lies on the fact that we assume complete ignorance regarding the possible extensions of the votes.

The results presented in [CLMM10] and subsequently extended in [CLM+10] focus on *positional scoring rules*. There are rules for which one can be rapidly convinced that a simple algorithm will do the job. Let us consider plurality. Intuitively:

- each new candidate can be placed on top of the votes to decrease the score of some candidate;
- for each candidate with a higher score than x^* we must put the new candidate on top a number of times equal to the difference of scores (+1 if that candidate has priority in the tie-breaking rule);
- the score of the new candidate must not be higher (or indeed equal if the new candidate has priority) than the current score of x^*.

This generalizes to k new candidates.

$$score_{Pl}(P_X, x^*) \geq \frac{1}{k} \sum_{z \in X} max(0, score_{Pl}(P_X, z) - score_{Pl}(P_X, x^*))$$

This is illustrated in the following example where the initial plurality scores are $s(a) = 3, s(b) = 2, s(c) = 2, s(d) = 1$. Certainly a (the current winner), and y (the new candidate) are possible winners. But what about the others? We see that applying the simple algorithm proposed above, b can be turned into a possible winner, but c cannot (y would then get a better score than him). In fact there are no wiser manner the cast the votes, so the condition stated above exactly characterizes when a candidate x^* is a possible winner upon arrival of k new candidates.

In [CLMM10], we identified several cases solvable by efficient algorithms (notably Borda, and K-approval with only one new candidate added). It turned out that the algorithm used

1: $a \succ d \succ c \succ b$	1: $a \succ d \succ c \succ b \succ \underline{y}$	1: $a \succ d \succ c \succ b \succ \underline{y}$
2: $a \succ b \succ c \succ d$	2: $\underline{y} \succ a \succ b \succ c \succ d$	2: $\underline{y} \succ a \succ b \succ c \succ d$
3: $a \succ d \succ c \succ b$	3: $\underline{y} \succ a \succ d \succ c \succ b$	3: $\underline{y} \succ a \succ d \succ c \succ b$
4: $d \succ a \succ c \succ b$	4: $d \succ a \succ c \succ b \succ \underline{y}$	4: $d \succ a \succ c \succ b \succ \underline{y}$
5: $b \succ a \succ c \succ d$	5: $b \succ a \succ c \succ d \succ \underline{y}$	5: $\underline{y} \succ b \succ a \succ c \succ d$
6: $b \succ d \succ a \succ c$	6: $b \succ d \succ a \succ c \succ \underline{y}$	6: $b \succ d \succ a \succ c \succ \underline{y}$
7: $c \succ d \succ a \succ b$	7: $c \succ d \succ a \succ b \succ \underline{y}$	7: $c \succ d \succ a \succ b \succ \underline{y}$
8: $c \succ b \succ d \succ a$	8: $c \succ b \succ d \succ a \succ \underline{y}$	8: $c \succ b \succ d \succ a \succ \underline{y}$

for the Borda rule can be used for a larger class of scoring rules, namely those where the scoring vector is concave. Intuitively, it corresponds to rules where the difference of points assigned to a position and the next one becomes smaller as we approach the top ranks. We also showed that K-approval could be hard, even for a small K and a small number of new candidates (specifically, as soon as $K \geq 3$ and $k \geq 3$), and perhaps more surprisingly that there exists a scoring rule for which the problem is hard even with a single new candidate. This rule is not a "common" voting rule, however it remains simple enough to be arguably used in real situations: it is a mixture of 3-approval and Borda, which assigns 3 points to the first candidate, 2 to the second one, 1 to the third one, and 0 to all the others. These results were further refined with the help of Lirong Xia in [CLM+10] to offer a complete picture of this problem, solving in particular the case of 2-approval left open in [CLMM10]. It is interesting to compare these results with the general version of the problem, as well as with other relevant problems such as the complexity of coalitional manipulation (see Table 2.4). Recently, the case of (some) non positional scoring rules (approval, Copeland, and Maximin)

Voting rule	*General problem*	*Missing candidates*	*Manipulation*
Plurality and Veto	P	P	P
Borda	NP-complete	P	open
2-approval	NP-complete	P	P
K-approval ($K \geq 3$)	NP-complete	NP-complete	P

has also been addressed in [XLM10].

A natural follow-up to this work would consist in considering elicitation protocols that can be designed, given that all the voters have expressed their vote, that these votes have been compiled using the techniques described in Section 2.2, and that some new candidates show up. In this case, can we use the stored information and do better than simply elicitate the preferences on the new set of candidates? To illustrate the idea, consider again the plurality rule, and suppose the profile was compiled anonymously by storing the plurality score of each candidate. Now a new candidate y is announced. The following protocol is possible:

(i) for each candidate, ask every voter whether they would now vote for y,

(ii) if the number of voters who said 'yes' is higher than n/2,
then y is the winner

else ask voters who said 'yes' the name of the candidate they ranked first before.

The communication complexity of this protocol is simple to compute: during step (i) we need n bits to get the yes/no replies of each voter, then during step (ii) we need to ask each voter who said 'yes' in the first phase the identity of their previous top-ranked candidate ($\log p$). But there are certainly at most $n/2$ such voters, otherwise the protocol would have stopped and declared y winner. This yields an overall commucation complexity of $n + \frac{n}{2} \log p$. As usual, proving that this is the best possible protocol (that is, the one that minimizes the number of bits exchanged in the worst case) is more challenging. It is often necessary to use techniques to prove lower bounds borrowed from communication complexity [KN97]. In this case, perhaps surprisingly, a better protocol exists...

Another difficult question that is raised by the voting applications discussed in this chapter is whether voters and/or candidates would happily accept the outcome of the process (of course this problem is only meaningful if this is a human being, or at least if there is a human being behind the agent candidate or voter). This is already not obvious in a classical setting, but looks even more critical here. Think of the 10% who were still to cast their vote when the voting procedure was declared closed. Think of a candidate eliminated whereas all the other ones can still hope to win the election. In this case it becomes very important to properly justify the decision taken. It is likely that an explicit proof (showing for instance that a candidate is no longer a possible winner under the Copeland rule...) would be to complex to handle and understand by a non-specialist user. This means that there is a need for specific arguments, tailored for this context. This topic is closely related to that of providing convincing arguments to a decision-maker regarding a recommendation given in the context of multi-criteria aiding [OMT07, OMT08]. We have started the investigation of this problem (for a specific case of decision procedure) with Alexis Tsoukiàs and Wassila Ouerdane, during her PhD thesis. The approach defines in particular a set of basic statements that can used to compose explanations. We are currently working in collaboration with Christophe Labreuche to connect this work to a related proposal of his own. This could lead to developments that may be of interest for the voting applications of this chapter. However, note as a final remark that although it is often tempting to equate criteria and voters, the contexts are not exactly similar (for instance, in multi-criteria decision-aiding, the information available to build explanations is less restricted since there a typically no concerns of anonymity for criteria).

3 Allocating Resources

Properties of systems where agents can reallocate resources among them by means of individually acceptable deals.

Resource allocation is a central topic in computer science and economics. The work we are going to present in this chapter is dedicated to a framework which, despite being sufficiently general to cater for a large variety of problems as we discuss in [CEEM04], makes some important assumptions. Firstly, the resources considered are *non shareable* and *indivisible*. This latter point excludes for instance the vast literature concerned with *cake-cutting* procedures [BT96].

The interest of computer scientists in resource allocation problems has been reinforced in recent years by the deployment of large-scale distributed applications involving a (sometimes very large) number of autonomous entities, possible synergies between resources, and where (optimal) central computation is in practice infeasible. Two canonical examples can be given: the coordination of task-allocation within a team of autonomous robots (multi-robot task allocation problems) [KTL+06], and the allocation of computational resources on grid-like systems [GP05]. All these systems typically build on the Contract-Net [Smi80], the famous protocol designed to assign a set of tasks to agents: from an initial allocation, agents reassign the tasks among themselves 'in some way', until a 'satisfying' allocation is found. As noticed in [SLB09], 'in some way' can be instantiated by a number of reallocation mechanisms. For instance, the multi-robot task-allocation problem has been successfully addressed in recent years by relying on a central auction mechanism computationally less demanding than a *combinatorial auction* [CSS06], but relaxing the optimality requirement: the sequential single-item auction proposed in [KTL+06] is provably close to the optimal. However such a solution still relies on central computation, which is arguably not always possible or desirable (as for instance in grid systems). In these cases a fully distributed approach is well-suited: a (typically small) subset of agents agree on (local) re-allocation of goods among themselves, and the system evolves as a consequence of these deals until a stable state is reached. Agents may be truly self-interested so these deals should be individually rational. This approach is in line with the *market-based programming* paradigm [Wel96].

In such applications, the designer of the system faces difficult problems: depending on the domain considered, protocol need to be finely tuned so as to ensure the properties sought (for

instance, convergence to a fair state) while remaining practically implementable. My various contributions to this area of research are in this line: the ambition is to provide insights to the designer as to what complexity of exchanges or additional payment schemes will need to be implemented to meet the announced objectives. Sometimes unfortunately we are just the bearer of bad news informing of the impossibility to guarantee a given property to hold.

This work was initiated as I was a postdoc researcher at Imperial College and City University (London), and was pursued in my current laboratory (LAMSADE). The results presented here are extracted from different papers. All of them are co-authored with Ulle Endriss, many of them with Yann Chevaleyre, and some of them with Sylvia Estivie, Francesca Toni, Fariba Sadri, and Jérôme Lang. In particular, fairness issues have been studied in the context of the PhD of Sylvia Estivie. This line of work also benefited from the support of different projects, which allowed to organize a series of informal "Multiagent Resource Allocation" meetings. The survey [CDE+06] written with several colleagues working on the subject results from one of these meetings, and is a good starting point to get a general overview of the field[1]. In the next section we provide the minimal necessary background on this approach. We then discuss in turn problems of convergence to efficient states, problems of convergence to fair states, and we briefly evoke a variant of the framework where no side-payments are involved, as well as the communication complexity of the process.

3.1 Background: A Distributed Resource Allocation Framework

3.1.1 Allocations

Let $\mathcal{G}$ be a finite set of indivisible *goods*. An *allocation* $A : \mathcal{N} \to 2^{\mathcal{G}}$ is a partitioning of the items in $\mathcal{G}$ amongst the agents in $\mathcal{N}$ (*i.e.* each good must be owned by exactly one agent). As an example, allocation A, defined via $A(i) = \{g_1\}$ and $A(j) = \{g_2, g_3\}$, would allocate g_1 to agent i, and g_2 and g_3 to agent j. There are $\mathcal{N}^{\mathcal{G}}$ possible allocations, and the system is initially in one of these allocations (that is, the goods are initially distributed). Observe that we do not put any constraints on the number of goods that agents may obtain, so that we have to deal with *bundles* of resources. If each agent could only hold a single good, the problem would be an *assignment problem*.

3.1.2 Preferences and Domains

Preferences of individual agents $i \in \mathcal{N}$ are modelled using *valuation functions* $v_i : 2^{\mathcal{G}} \to \mathbb{R}$, mapping bundles of goods to the reals. We shall assume that agents only care about the bundle they actually hold (*no externality*), so we can safely use $v_i(A)$ as a shorthand for $v_i(A(i))$, the value agent i assigns to the bundle received in allocation A. We also make the assumption that all valuations v_i are *normalised*, in the sense that $v_i(\{\}) = 0$. Sometimes the scenario takes place in a specific *domain*, in the sense that all agents' valuations functions are known to belong to restricted class of functions. The most natural domains are as follows. For all bundles $S_1, S_2 \subseteq \mathcal{G}$:

- *monotonic*: $S_1 \subseteq S_2$ implies $v_i(S_1) \leq v_i(S_2)$.

[1] For slight variants of this problem we refer *e.g.* to [AE10, SS07, BR08]

- *modular* (or *additive*): $v(S_1 \cup S_2) = v(S_1) + v(S_2) - v(S_1 \cap S_2)$
- *supermodular*: $v(S_1 \cup S_2) \geq v(S_1) + v(S_2) - v(S_1 \cap S_2)$.
- *k-additive* if it can be written as a sum of *terms*

$$v_i(S) \quad = \quad \sum_{T \subseteq \mathcal{G},\, |T| \leq k} \alpha_i^T \times I_R(T) \quad \text{with } I_R(T) = \begin{cases} 1 & \text{if } T \subseteq S \\ 0 & \text{otherwise} \end{cases}$$

- *tree-structured* if, when written under the k-additive form, it is the case that for all terms $T_1, T_2 \in \mathcal{T}$ (the set of non-null terms of the function) we have either $T_1 \subseteq T_2$, or $T_2 \subseteq T_1$, or $T_2 \cap T_1 = \{\}$.
- additively *k-separable* wrt. a partition $P = \langle R_1, \ldots, R_q \rangle$ of $\mathcal{G}$, if the following holds
  (i) $|R_j| \leq k$ for all $j \in \{1..q\}$, and
  (ii) for all $S \subseteq \mathcal{G}$:

$$u(S) = u(\{\}) + \sum_{j=1}^{q} [u(S \cap R_j) - u(\{\})]$$

- *dichotomous* if $v(S) = 1$ or $v(S) = 0$ for all $S \subseteq \mathcal{G}$.

In words, monotonicity says that agents are always as least as happy with a superset of their resources. This will be assumed throughout this chapter. Modularity means that no synergy occurs between resources. Super-modularity (resp. sub-modularity) allows only positive (resp. negative) synergies. With k-additivity [Gra97], both synergies are allowed, but only up to subsets of cardinality (at most) k. Clearly, k-separability is more demanding: it requires synergies between items to be limited to specific subsets. For instance, we preparing a picnic basket, perhaps wine and cheese have synergies, but fruits can be considered independently. Finally, tree-structured domains forbid terms of utility functions (when expressed in the k-additive form) to be overlapping. Note that additive k-separability implies k-additivity. The converse is not true of course, however a tree-structured k-additive domain is also additive k-separable. When $k = 1$ both notions simply correspond to modularity. It is important to bear in mind that when we refer to an additive k-separable domain, the partition has to be common among all agents. Similarly, a tree-structured domain requires non-overlapping terms among all agents utility functions.

3.1.3 Representing Preferences over Bundles

As we said, the domain over which the decision must be made is inherently combinatorial, because agents negotiate over the possible allocations. Even if each agent is only concerned with its own bundle, an explicit representation consisting of enumerating the 2^m bundles is infeasible for a moderate number of resources. In consequence, a question arises: how to *concisely* represent preferences? Representation languages may be compared on the basis of their *conciseness*, and of their *expressivity* [Nis06]. There are several proposals available, for instance *bidding languages* [Nis06] developed in the context of combinatorial auctions, *logic-based* approaches [LL00, UCEL09], or the recent CI-net [BEL09] proposal, which allows

conditional statements (in the spirit of CP-nets [BBD+04] but adapted to the resource allocation contexts). The issue is important in particular when we consider the computational complexity of decision problems, because the representation of some preference structure may be exponentially larger than others in some cases. At this point we do not commit to any specific representation language, but will mention the matter when necessary.

For practical matters, the k-additive representation is often used in this chapter (it is fully expressive when no restriction is put on the value k). As an example, take the following utility function, representing in (slightly simplified) k-additive form the preferences of agent a_1 over the possible bundles composed from $\{g_1, g_2, g_3\}$.

$$u_1 = 3g_1 + 5g_2 + 2g1_g3 + 2g_2g_3$$

The interpretation is as follows: for instance, if a_1 obtains g_3 alone her utility is 0, if she gets $\{g_1, g_2\}$ her utility is 8, if she gets $\{g_1, g_2, g_3\}$ her utility is 12. The utility is 2-additive, not tree-structured (because the terms g_1g_3 and g_2g_3 are overlapping), and not 2-separable (but obviously 3-separable, since there are only 3 resources).

3.1.4 Side Payments

An important distinctive feature of resource allocation frameworks is whether or not they allow to use *money*. Money may be used in deals to compensate the utility loss of some agents. Money also makes more acceptable interpersonal utility comparisons. Most of the work described here is in a framework *with* money (we shall nevertheless provide some insights on the framework without money in Section 3.4), meaning that deals may be accompanied by monetary side payments. This is modelled using so-called *payment functions:* $p : \mathcal{N} \to \mathbb{R}$, which are required to satisfy $\sum_i p(i) = 0$. A positive value $p(i)$ indicates that agent i *pays* money, while a negative value means that the agent *receives* money. We associate each allocation A that is reached in a sequence of deals with a function $\pi : \mathcal{N} \to \mathbb{R}$ mapping agents to the sum of payments they have made so far, *i.e.* we also have $\sum_i \pi(i) = 0$. It is possible to impose an *initial payment* on each agent, at the time of awarding them the bundle they receive in the initial allocation. Payment functions and initial payments together are referred to as the *payment scheme*. In this context, a *state* of the multiagent system is described as the current allocation of resources, together with a *payment balance* π. Each agent $i \in \mathcal{N}$ is then equipped with a *utility function* $u_i : 2^{\mathcal{G}} \times \mathbb{R} \to \mathbb{R}$ mapping pairs of bundles and previous payments to the reals. These are fully determined by the valuation functions: $u_i(S, x) = v_i(S) - x$. That is, utilities are *quasi-linear:* they are linear in the monetary component, but the valuation over bundles of goods may be any (monotonic) set function. For example, $u_i(A(i), \pi(i))$ is the utility of agent i in state (A, π).

3.1.5 Deals and Individual Rationality

Very abstractly, a *deal* can be described as a pair of (distinct) allocations $\delta = (A, A')$, fixing the situation before and after the exchange. When no restriction applies we sometimes talk of *complex deals*. In practice, the range of feasible deals are restricted, by limiting the number of resources or agents involved, or by preventing agents from giving and receiving resources at the same time. We also consider the case of topological restrictions affecting possible exchanges. Specifically, we assume a topological structure $G = (\mathcal{N}, E)$, an undirected graph specifying which agents can interact with each others. By default, a fully connected graph

is considered. Furthermore, the following types of deal restrictions are considered in this chapter[2] considered:

- *one-resource-at-a-time* (or *1-deals*): a single resource is passed from an agent to another;
- *swap deals*: one agent gives a resource and receives (simultaneously) a resource in return from another agent;
- *cluster deals*: a bundle of resources is passed from an agent to another agent;
- *bilateral deals*: only *two* agents are concerned (all previous cases are special cases of bilateral deals);
- $\mathcal{T}$*-deals*: must involve an entire term (from a set of terms $\mathcal{T}$), from one or more sender(s) to a single receiver;
- *clique deals*: involves only agents belonging to a common *clique*[3] of the graph G.

An agent may or may not find a particular deal $\delta = (A, A')$ *acceptable*. Agents are assumed to negotiate *individually rational* (IR) deals, *i.e.* deals that benefit everyone involved:

Definition 3.1.1 (IR deals) *A deal $\delta = (A, A')$ is called individually rational (IR) if there exists a payment function p such that $v_i(A') - v_i(A) > p(i)$ for all agents $i \in \mathcal{N}$, except possibly $p(i) = 0$ for agents i with $A(i) = A'(i)$.*

It is important to stress that this definition of individual rationality (i) is *myopic*, in the sense that agents do not plan ahead (for instance speculate by acquiring a resource they hope to sell for higher price at a later stage of the negotiation), and (ii) is *local* in the sense that each agent can individually decide or not whether the deal is acceptable from his point of view. Considering notions of *farsighted rationality* [Kaw10, KKW10] is a promising line of research for the future.

3.1.6 Efficiency and Fairness of Allocations

We are interested in reaching states that are attractive from a "social" point of view [Mou88]. As already mentioned, we may consider issues of efficiency and fairness. Now we make these notions more concrete.

Efficiency. When money is not available in the system, the notion of efficiency we rely on is that of Pareto-efficiency, unless stated otherwise. When interpersonal comparison of preferences intensities are meaningful, a common metric for efficiency is the *utilitarian* social welfare:

Definition 3.1.2 (Social welfare) *The utilitarian social welfare of an allocation A is defined as $sw(A) = \sum_{i \in \mathcal{N}} v_i(A(i))$.*

In our context, we speak of the social welfare of a *state* (A, π). As the sum of all $\pi(i)$ is always 0, it does not matter whether we define social welfare in terms of valuations or in terms of utilities.

[2] This list extends over a classification initially provided by Sandholm in [San98].
[3] A clique is a set of vertices $C \subseteq \mathcal{A}$ such that $(i, j) \in E$ for all distinct $i, j \in C$

Fairness. In terms of *fairness*, we shall mainly consider two metrics here. The first one, the *egalitarian* social welfare [Raw71], looks at the welfare of the poorest agent of the system (an tries to maximize it). The second one, *envy-freeness* [FK74], inspects whether (some) agents think they would be better off with the bundle of some other agents.

Definition 3.1.3 (Egalitarian social welfare) *The egalitarian social welfare $sw_e(A)$ of an allocation of resources A is defined as follows:*

$$sw_e(A) \quad = \quad \min\{u_i(A) \mid i \in \mathcal{A}\}$$

Definition 3.1.4 (Envy-freeness) *A negotiation state (A, π) is called envy-free if $u_i(A(i), \pi(i)) \geq u_i(A(j), \pi(j))$ for all agents $i, j \in \mathcal{N}$.*

Note that envy-freeness is a yes-or-no notion, but that it is possible to come up with various meaningful quantitative measures of envy (for instance by counting the number of envious agents, etc.), as we discuss in [CEM10a].

The Efficiency *vs.* Fairness Trade-Off. Usually both criteria cannot be optimized for the same state, so a trade-off between efficiency and fairness is in order. However a designer may want to design the system so as to reach the fairest among potentially many efficient allocations, or alternatively, the most efficient among equally fair allocations. It is also important to note that the possibility to use payments is highly helpful, because redistributing this continuous resource makes precise compensations easy. In this context the trade-off often vanishes. Take egalitarian social welfare. In the presence of money, the optimal solution is trivial: it suffices to compute the optimal utilitarian allocation, and then design the payment so as to make everyone exactly equally happy. Another case of special interest is envy-freeness. In the presence of money there is *no* trade-off either, that is, states that are both efficient (in the sense of maximising utilitarian social welfare) and envy-free are guaranteed[4] to exist [ADG91, Bev98] (these states will be referred to as *EEF states*). This is not true when money is not allowed in the system, and actually identifying whether some states are both Pareto-efficient and envy-free is far from trivial [BL08, dKBKZ09]. There is another interesting underlying question here, namely: what is the "price of fairness", that is, what is the loss of efficiency if you enforce a solution to be fair (in some of the senses defined earlier). The problem is studied in [CKKK09].

Computational Complexity. The reader will not be suprised to learn that in most cases, solving centrally the decision version of these optimization problems is intractable. In particular, finding an optimal allocation in terms of utilitarian social welfare precisely correspond to the *winner determination problem* in combinatorial auctions and is known to be NP-complete [RPH98] , even for non concise languages like XOR [LMS06]. In general, fairness involves measures that appear to be harder to optimize than those of efficiency: in fact, their intractability does not necessarily come from the combinatorial nature of the domain, whereas for instance optimizing utilitarian social welfare becomes easy in modular domains[5]. Some

[4] More precisely, Bevià proves that with quasi-linear utilities [Bev98], the existence of envy-free allocations can be guaranteed.

[5] As a very simple example of this, the maxmin optimisation induced by the egalitarian problem is hard, in the absence of money, even in modular domains.

of these problems lie beyond NP, so even verifying that a solution may not be feasible in polynomial time. Although not directly touching the distributed feature, these results are of course highly relevant for this approach.

3.1.7 Specific payment functions

A key result [EMST03a] of the framework is that a deal (with money) is individually rational *iff* it increases utilitarian social welfare. This means in particular that any IR deal $\delta = (A, A')$ generates a *social surplus* $(sw(A') - sw(A))$. And this in turn raises the question of choosing a payment function, that is, choosing how to distribute this *social surplus* generated by the deal. There are several options to do that. We start with, arguably, the most natural ones.

- the *locally uniform* payment function (LUPF) divides this amount equally amongst the *participating* agents $\mathcal{N}^\delta$ (and does not affect the other agents);

$$p(i) = [v_i(A') - v_i(A)] - [sw(A') - sw(A)]/|\mathcal{N}^\delta$$

- the *globally uniform* payment function (GUPF) divides it equally amongst *all* agents $\mathcal{N}$:

$$p(i) = [v_i(A') - v_i(A)] - [sw(A') - sw(A)]/n$$

These payment functions are "uniform" in the sense that they redistribute the surplus, without any consideration for agents' current situation. We may want to design payment functions that perform some compensation, in the sense that agents are that worst-off receive more from the social surplus generated. In [ECEM06] we introduced the following payments:

- the *fully locally equitable* payment function, arrange the payment such that each agent (involved in the deal) would enjoy the same utility level after the deal has been achieved. An important aspect of this payment function is that it may enter in conflict with IR.
- the *rational locally equitable* payment function is computed so as to make every agent marginally better off (so as to satisfy IR), then allocate the remaining payments induced by the deal so as to reduce inequalities as much as possible.

3.2 Convergence to Efficient Outcomes

In this section we consider the following general question: will the system converge to an efficient state, if agents autonomously agree on deals satisfying their individual rationality criterion? A central result in this distributed negotiation setting is due to Sandholm:

Theorem 3.2.1 (Efficient outcomes [San98]) *Any sequence of IR deals will eventually result in an allocation maximizing utilitarian social welfare.*

This theorem is a result of *guaranteed convergence*: it guarantees that, starting from *from any initial allocation*, agents can never get stuck in a local as long as they implement IR deals. Unfortunately, it is also known that for such a positive result to hold, one has to allow very complex deals to be implemented (involving arbitrary many resources and agents). We showed that this remains true even if agents' valuations are monotonic, or dichotomous:

Theorem 3.2.2 (Necessary deals with side payments [EMST06]) *Let the sets of agents and resources be fixed. Then for every deal δ, there exist utility functions and an initial allocation such that any sequence of individually rational deals leading to an allocation with maximal utilitarian social welfare would have to include δ.*

Some seemingly very severe restrictions are not more helpful. The following example, taken from [CEEM08], shows that focusing on k-additive domains is also unlikely to allow simpler protocols to be used. The deals required to reach maximal social welfare in the k-additive case remain very complex and require combined deals.

To show this, let us build an example with 2-additive utility functions. Consider 2 agents sharing n resources $\{r_1, r_2, ...r_n\}$, with the following 2-additive utility functions : $u_1 = 0$ and $u_2 = r_1 - r_1.r_2 - r_1.r_3 - r_1.r_4 - \ldots - r_1.r_n$. In words, agent a_1 is only interested in resource r_1, but this resource has a negative synergy with *all* other resources. Let A_{init} be the initial allocation describing which agent owns which resource at time 0, and let A_{opt} be the allocation maximizing the utilitarian social welfare.

	A_{init}	A_{opt}
a_1	$\{r_1\}$	$\{r_2, r_3, \ldots, r_n\}$
a_2	$\{r_2, r_3, \ldots, r_n\}$	$\{r_1\}$

Here, $sw(A_{init}) = 0$ and $sw(A_{opt}) = 1$. In fact, the *only* allocation which has a social welfare greater than $sw(A_{init})$ is A_{opt}. which is a bilateral deal of n resources at a time. The only IR deal is $\delta(A_{init}, A_{opt})$. It consists in getting rid of all resources, and obtaining r_1 in return, a combined deal involving all n resources.

3.2.1 Sufficient Domains

Fortunately, if we sufficiently restrict the domain from which agents' preferences can be drawn, it becomes possible to guarantee convergence despite protocols that allow only restricted type of deals. The simplest example is that of *modular domains*, and can be stated like that :

Theorem 3.2.3 (Sufficiency of 1-deals in modular domains [EMST06]) *In modular domains, any sequence of IR 1-deals will eventually result in an allocation maximizing utilitarian social welfare.*

In a series of companion papers we provide more results of the same flavour, investigating domains that guarantee convergence, whether or not side payments are allowed. In [CELM05], we sought to characterize the adequate domain, when the negotiation protocol allows k resources to be passed from one agent to another agent. It turns out that in this case k-separable domains do the job.

A problem with this is that the number k-deals may already be too large to consider. In [CEM06] we studied how the property of tree-structure (when the domain satisfies it[6]) may be exploited. In this case, the idea is to let agents negotiate via $\mathcal{T}$*-deals* only, for indeed the number of possible $\mathcal{T}$-deals is low ($n \times |\mathcal{T}|$). The bad news is that simply allowing any $\mathcal{T}$-deals does not guarantee convergence to optimal utilitarian social welfare. This is illustrated by

[6]Recall that k-additive tree-structured utility functions are k-separable, so this is a stronger requirement than k-separability.

the following example: suppose $u_1 = 10.g1$,$u_2 = 10.g_2$, and $u_3 = 15.g_1g_2$, and that initially all resources are held by agent 3. The only IR deal requires a_3 to simultaneously give g_1 (to a_1) and (to a_2), but this is not a $\mathcal{T} - deal$.

The proposed solution is to to introduce an element of centralization (the *bank* agent), in order to guide the negotiation process in the following sense: (1) deals of higher complexity will be incrementally allowed by the system, (2) appropriate payments, involving the bank agent, will be implemented in such a way that negotiation does not end in a suboptimal state. The resulting negotiation protocol, called *Omniscient ϵ-Altruistic Tree-Climbing*, provably converges to maximal utilitarian social welfare. The parameter ϵ affects the way payment function redistribute the surplus, making the bank agent more or less altruistic. In this case, the convergence property is guaranteed at the price of a severe limitation of the decentralized view advocated here, since agents need in particular to trust the bank agent and reveal their utility functions to allow the correct implementation of the payment scheme involved in the protocol.

3.2.2 Lack of Necessity

Perhaps the domains identified by our sufficiency results are required to make convergence a guaranteed property. Take convergence by means of 1-deals: maybe modularity exactly characterizes those domains guaranteeing the property? After a moment thought, one sees although that this is obviously not the case: there are plenty of domains where the property is also ensured. Sometimes these domains are rather silly, like the *pseudo-constant* domain where agents equally like any allocation where they get at least *some* resource, whatever the resource(s). In fact it is more than that. It is possible to show that there can be no domain of valuation functions that would be both sufficient and necessary. Suppose a necessary and sufficient domain (say, $\mathcal{D}$) exists. Let us take v_1 as being a modular function. Certainly if all agents are using v_1 convergence is guaranteed, so v_1 must belong to $\mathcal{D}$. Let us now take v_2 as being a pseudo-constant valuation. Here again, if all agents are using v_2 convergence is guaranteed, so v_2 must also belong to $\mathcal{D}$. As a consequence any scenario involving agents using either v_1 or v_2 must also necessarily converge. The counter-example provided in Table 3.1 ruins our best hopes to ever find such a domain. It involves two agents and two resources (the argument is easily augmented to the general case).

$u_1(\{\,\}) =$	0	$u_2(\{\,\}) =$	0
$u_1(\{\spadesuit\}) =$	4	$u_2(\{\spadesuit\}) =$	$\boxed{1}$
$u_1(\{\clubsuit\}) =$	$\boxed{4}$	$u_2(\{\clubsuit\}) =$	3
$u_1(\{\spadesuit,\clubsuit\}) =$	4	$u_2(\{\spadesuit,\clubsuit\}) =$	4

$u_1(\{\,\}) =$	0	$u_2(\{\,\}) =$	0
$u_1(\{\spadesuit\}) =$	$\boxed{4}$	$u_2(\{\spadesuit\}) =$	1
$u_1(\{\clubsuit\}) =$	4	$u_2(\{\clubsuit\}) =$	$\boxed{3}$
$u_1(\{\spadesuit,\clubsuit\}) =$	4	$u_2(\{\spadesuit,\clubsuit\}) =$	4

Table 3.1: A scenario involving two agents where a swap deal is necessary

Preferences of agent a_1 are pseudo-constant, while that of agent a_2 are modular. The table on the left is the initial allocation, yielding a utilitarian social welfare of 5. The table on the right shows the optimal allocation (the social welfare is 7). The reader can easily check that no sequence of IR 1-deals possibly leads to this allocation (a swap deal would be required here).

3.2.3 Maximal domains

Given the previous findings on the non-existence of domains exactly characterizing guaranteed convergence, the next big thing would be to identify *maximal* domains exhibiting this property. A domain is said to be maximal when *any* larger domain (strictly including it) loses the property of guaranteeing convergence to maximal utilitarian social welfare. With Yann Chevaleyre and Ulle Endriss we introduced the question and proved maximality of the modular domain for 1-deals (even for scenario involving two agents) in [CEM05]. In [CEM10b] we were able to prove that the modular domain is maximal for guaranteed convergence by means of the much larger class of bilateral deals.

Theorem 3.2.4 (Maximality wrt. bilateral deals [CEM10b]) *Let $\mathcal{M}$ be the class of modular valuation functions. Then for any class of valuation functions $\mathcal{F}$ such that $\mathcal{M} \subset \mathcal{F}$, there are negotiation scenarios with valuation functions drawn from $\mathcal{F}$ such that no sequence of IR bilateral deals will lead to an allocation with maximal social welfare.*

To prove this kind of results, we proceed by constructing a situation such that, (i) for an arbitrary agent's utility function *not* picked from the domain, we can construct a scenario where (ii) all the other agents' utility functions are, and such that (iii) from a given initial allocation no sequence of deals (respecting some constraints) can lead to the optimal outcome. This suffices to show failure of convergence, since this property should hold regardless of the initial state. The question is studied in detail in [CEM10b].

Let us stop a second on the practical consequences of such results: what this means is that a designer implementing a multiagent system where agents can only interact by means of bilateral deals can only[7] hope to guarantee convergence as long as each agent indeed exhibits a modular utility function. This does not rule out that for a specific scenario, convergence may be guaranteed. So, is it really a problem after all? The designer may instead check whether the scenario at hand guarantees convergence. There are two important problems with that:

- this requires the designer to know exactly the full profile, *i.e.* the different preferences of all agents involved in the system (as opposed to just knowing that agents' preferences are drawn from a specific domain).
- the related decision problem is intractable for most representation languages, at least those sufficiently compacts[8] [CEM10b].

The complexity study of multiagent resource allocation problems has been initiated by Paul Dunne an his co-authors. In [DWL05], they address the following question: what is the complexity of deciding whether there exists a sequence of deals leading to given allocation, such that all deals are restricted in terms of the number of resources that can be transmitted, and such that all deals in the sequence provide an immediate welfare improvement to the agents involved? The results presented conclude on the intractability of these decision problems. In fact, this result is strengthened in [DC06]: this reachability property turns out to be PSPACE-complete (so is unlikely to be in NP). This is the main result presented in

[7] In fact, there may be *other* maximal domains, but the modular domain is arguably one of the most natural one.

[8] This excludes for instance the bundle form which requires to list explicitly the valuations corresponding to all possible bundles.

this paper, a result that must be appreciated in contrast to the "mere" NP-completeness of optimally allocating the resources by means of a centralised mechanism, for instance combinatorial auctions.

3.3 Convergence to Fair Outcomes

We now investigate whether it is possible to design our mechanism so as to guarantee that the allocation of resources that will be reached eventually will be fair. In what follows we discuss the case of egalitarian social welfare, as well as envy-freeness. In [EM04, EMST06] will also give some results regarding the convergence to *Lorenz*-optimal outcomes, whereas in [CEM10a] the case of *proportionality* is discussed.

3.3.1 Convergence to Egalitarian Outcomes

A first natural approach is to modify the acceptability criterion of agents, in such a way that their local decisions reflect the global measure we seek to optimize. The notion of equitable deals follows this idea.

Definition 3.3.1 (Equitable deals [EMST03c]) *A deal $\delta = (A, A')$ is called equitable iff it satisfies the following criterion:*

$$\min\{u_i(A) \mid i \in \mathcal{A}^\delta\} \quad < \quad \min\{u_i(A') \mid i \in \mathcal{A}^\delta\}$$

Recall that $\mathcal{A}^\delta = \{i \in \mathcal{A} \mid A(i) \neq A'(i)\}$ denotes the set of agents involved in the deal δ. Given that for $\delta = (A, A')$ to be a deal we require $A \neq A'$, $\mathcal{A}^\delta$ can never be the empty set (the minima referred to in above definition are well-defined). The careful reader will have noticed something unsatisfying with this definition: it violates the requirement of locality that we mentioned earlier. Indeed, agents would have to communicate in order to check that the poorest among them is really better-off after the deal. Furthermore, if we are only concerned with the poorest agent of the society, then this criteria looks too "strong": there are situations where there would still be some deals possible, whereas the outcome is optimal already.

Prompted by these remarks, we investigated in [ECEM06] whether the IR criterion would lead in practice to egalitarian outcomes. As mentioned before, in that case the outcome is very much sensible to what specific payment function is used to distribute the social surplus generated by a specific deal. In [ECEM06] we study the respective merits of some of the payment functions introduced in sub-section 3.1.7. A number of experiments were performed, in both modular and non-modular domains, and we studied the outcomes reached at the end of the negotiations. Our main conclusions can be summarized as follows: unsurprisingly (1) the higher the number of agents, the more important it is to rely on a global payment function to compensate inequalities; and more surprisingly (2) in modular domains, the LUPF gives rise to very equitable outcomes, almost attaining the best we can hope for under the constraints of IR (in particular, it is almost on tie with outcomes obtained under the rational equitable payment).

3.3.2 Convergence to Envy-Free Outcomes

Intuitively, convergence to envy-free outcomes seems even more challenging to attain. A very simple example can for instance convince the reader that envy-freeness and IR are not

necessarily compatible. Take two agents and just a single good, such that $v_1(\{g\}) = 4$ and $v_2(\{g\}) = 7$ Suppose agent 1 holds g in the initial allocation A_0. There is only a single possible deal, which amounts to passing g to agent 2, and which will result in the efficient allocation A^*. How should payments be arranged? To ensure that the deal is IR for both agents, agent 2 should pay agent 1 any amount in the open interval $(4, 7)$. On the other hand, to ensure that the final state is envy-free, agent 2 should pay any amount in the closed interval $[2, 3.5]$. The two intervals do not overlap. This means that, while we will be able to reach negotiation outcomes that are EEF, it is simply not possible in all cases to do so by means of a process that is fully IR.

In [CEEM07], we proved that (efficient) envy-free outcomes can however be attained by any sequence of IR deals, provided that (1) we restrict our attention to supermodular domains, (2) we implement initial payments prior to the negotiation, and (3) we use the globally uniform payment function (GUPF). The result is formally stated as follows:

Theorem 3.3.1 (Envy-free outcomes [CEEM07]) *If all valuation functions are supermodular and if initial equitability payments have been made, then any sequence of IR deals using the GUPF will eventually result in an EEF state.*

A couple of remarks are required here regarding the use of the GUPF. It certainly adds a non-local element. However, note that only the agents involved in a deal can ever be asked to give away money, and all payments can be computed taking only the valuations of those involved agents into account.

Moreover, we have good reasons to believe that this about the best we can hope for, within the realms of this framework. Indeed, we were able to show in [CEM10a], much in the spirit of the maximality results mentioned in Section 3.2.3, that for any domain strictly larger than the supermodular domain, it would be possible to construct a situation showing failure of the convergence property.

3.3.3 Convergence to Envy-Free Outcomes on Graphs

So far the results mentioned made the implicit assumption that no topological restrictions apply on the possible interactions between agents. In fact, any restriction can severely damage efficiency in the worst-case: it suffices to construct a scenario where precisely a deal between two unconnected agents would be necessary to attain the efficient outcome. Interestingly however, the notion of envy-freeness accepts a variant where the underlying topology of the system is considered. Intuitively, an agent may only envy another agent if it sees it, *i.e.* if both agents are connected in the graph. We introduced in [CEM07] this notion of *graph envy-freeness*:

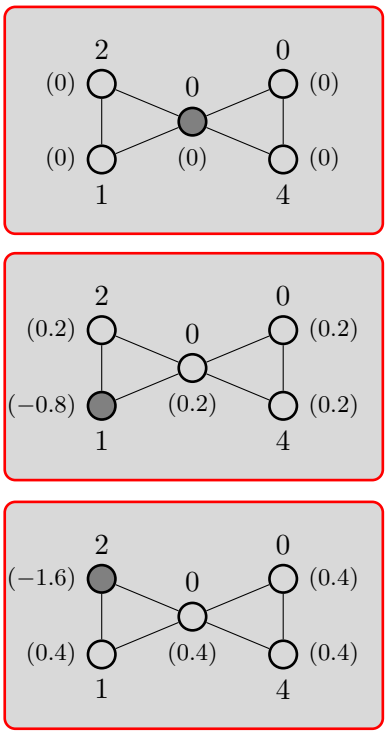
Definition 3.3.2 (GEF states) *A state (A, π) is called graph-envy-free (GEF) with respect to the graph $G = (\mathcal{N}, E)$ if $u_i(A(i), \pi(i)) \geq u_i(A(j), \pi(j))$ for all agents $(i, j) \in E$.*

Unfortunately, as we have seen before, convergence to envy-freeness relies on the final allocation being efficient (or at least efficient "enough"). It turns out that a notion of efficiency that takes the negotiation topology into account can be developed. And we will show that this notion will suffice to guarantee the graph envy-freeness property we want to see satisfied. Let us call a *clique-variant* a possible reallocation of goods within the agents of a given clique, in other words A and A' are clique-variants of each other iff $\delta = (A, A')$ is a clique-deal.

Definition 3.3.3 (Clique-wise efficiency) *An allocation A is called clique-wise efficient if $sw(A) \geq sw(A')$ for every clique-variant A' of A.*

The ingredients are in place to give a convergence theorem for GEF states, which extends Theorem 3.3.1 to the framework with a negotiation topology: as we show in [CEM07], under the same conditions (on the valuation functions and for a particular choice of payment scheme), any sequence of IR deals that respect the negotiation topology will result in a GEF state.

Theorem 3.3.2 (Convergence on graphs [CEM07]) *If all valuations are supermodular and if initial equitability payments have been made, then any sequence of IR clique-deals using the GUPF will eventually result in a GEF state.*



Initial allocation.
The middle agent holds the resource.
No initial payments are implemented.

The resource is bought by the bottom-left agent.
The deal generates a surplus of 1. The buyer has to pay $(1 - 0) - 0.2 = 0.8$. The seller pays $(0 - 0) - 0.2 = -0.2$ (gets 0.2), as the other agents. At this point, the top-left agent envies the bottom-left agent.

The resource is eventually bought by the top-left agent.
The deal generates again a social welfare surplus of 1.
The buyer pays $(2 - 0) - 0.2 = 1.8$. The seller gets 1.2 $(= (0 - 1) - 0.2)$. All the other agents get 0.2. No further deal is possible. Note that the top-right agent would be envious of the top-right agent if he could see her.

Figure 3.1: Convergence to graph envy-free states illustrated

Firgure 3.1 illustrates the convergence result. There are five agents (nodes), dispatched on a network composed of two cliques (the center agents belong to both cliques). One resource is to be allocated, and the respective utilities of the different agents are given next to the nodes. For instance, the top-left agent values the resource 2. The resource is initially held by the center agent, as indicated by the grey node. The payment balance of the different agents are given into brackets. A possible sequence of two deals leading to a clique-efficient allocation is given, and the details of the computation of the payments is given as comments to the figure.

3.4 The Framework without Money

All the results presented so far are in the variant of the framework where money is available to agents, in particular to add side-payments to deals. In many situations though, the use of money is not possible, for practical or ethical reasons. The same kind of analysis can be performed, and we give a quick overview of some of the results we managed to obtain in this context.

When no side-payment is allowed, one can see that acceptable deals will be much more difficult to implement. For instance, 1-deals would never be used (if valuations functions are non-negative), since the agent giving the resource doesn't get anything in return. More extremely, suppose an agent values any resource at 0, but get all the resources in the initial allocation. No deal would ever be possible, so Pareto-optimal allocations would not be reached, let alone allocation with maximal utilitarian social welfare. To account for that, the individual criteria of rationality of agents is relaxed, and we assume instead that they are *cooperative rational* [EMST06]. This means only one agent needs to experience a strict increase of utility, the other ones can accept the deal as long as their utility does not decrease. This mirrors at the local level the notion of Pareto improvement, and as hoped any sequence of cooperative rational deals will converge to a Pareto-optimal allocation of resources. As in the case with money, we were also able to obtain negative results regarding the potential necessity to use complex deals [EMST06], even when preferences are monotonic or dichotomous. Finally, we identified in [CEM10b] a sufficient and maximal domain for guaranteed convergence to optimal utilitarian social welfare. The careful reader might stop us here: didn't we argue before that the use of the utilitarian metric in a framework without money was dubious? True, but the domain we are about to introduce is precisely a domain that, in a sense, allows to circumvent this issue. The domain is a restriction on the class of modular valuation functions, namely the classes of so-called *modular functions with fixed α, β-values*. Intuitively, $\mathcal{M}_{\alpha,\beta}$-classes are suitable in cases where agents can only like, dislike, or possibly be indifferent towards any given resource in the system. The key point is that agents *all agree on the intensities used to indicate positive and negative preferences for each single resource*, so the problem of intercomparable utilities does not occur.

3.5 Number of Deals Required to Reach Outcomes

Once we know that a system is guaranteed to converge, we may wonder how quickly the system will reach its outcome. There may be other interpretations of what communication complexity means in that context [EM05], but we are going to address the following question: how many deals are required to reach an optimal allocation of resources?

The class of deals considered (with or without money; one-resource-at-a-time or general) as well as the type of optimality that can be achieved (maximal social welfare or Pareto optimality) differ for the different results presented next. For instance, we are going to investigate the number of rational 1-deals with money required to reach an allocation with maximal social welfare in a modular domain.

Of course, 0 is always going to be a *lower bound:* if the initial allocation of resources is itself optimal, then not a single deal will be required to reach an optimal allocation. Hence, we are only going to be interested in upper bounds. In fact, there are two types of upper bounds that one may consider: the maximal length of the *shortest path* to an optimal allocation and

the maximal length of the *longest path* to such an allocation. For instance, in the general domain, the upper bound on the shortest path is 1: there is a way to jump from any allocation to the optimal one, and if agents are wise enough they will find it. This is not the case when deals are constrained to be 1-deals: it may be necessary to implement more complex deals to reach the optimal. As for longest paths, we need to compute how many allocations can be visited at most before reaching an optimal allocation. Table 3.2 offers an overview of some results obtained in [EM05]. Note that the bound for the longest path without money is not tight, contrary to the other ones.

Domain	general	general	modular
Side payments?	yes	no	yes
Deal types	any	any	1-resource
Shortest path	1	1	$\lvert\mathcal{G}\rvert$
Longest path	$\lvert\mathcal{A}\rvert^{\lvert\mathcal{G}\rvert} - 1$	$< \lvert\mathcal{A}\rvert \cdot (2^{\lvert\mathcal{G}\rvert} - 1)$	$\lvert\mathcal{G}\rvert \cdot (\lvert\mathcal{N}\rvert - 1)$

Table 3.2: How many rational deals are required to reach an optimal allocation?

Dunne [Dun05] tackles the same question, restricting his attention to 1-deals. It is shown that it may be necessary to implement a sequence of reallocation contracts of exponential length, even if the utility functions of agents are monotonic, and for a wide range of acceptability criterion. I emphasize that these results are *lower bounds*, so they hold regardless of how wise or lucky are agents when visiting allocations: in some situations even the shortest sequence of contracts would have to be exponential.

In [CEM08], we sketch a theoretical analysis of the *average communication complexity* of our framework, focusing here again on the case of 1-deals. Specifically, we conduct this analysis in the case where an underlying topology affects the possible deals. In that case we touch a delicate issue, for it becomes important to consider agents' *strategies*, that is, how will agents select their trading partner when faced with several (rational) deal options in their neighbourhood. In fact it is possible to identify several classes of strategies, depending on the input required to compute the partner agent with whom the deal will be contracted.

- *blind* strategies—agents are only allowed to check which of their neighbours are proposing rational deals to make their decision. For instance, on the basis of this information, we can assume that the agent will select the first partner proposing a rational deal following predefined lexical order; or alternatively just pick one at random until a rational deal can be implemented.

- *heuristic* strategies—agents select their partner on the basis of heuristics regarding the potential utility profit among all the neighbours. In the context of myopic agents, a certainly natural strategy is to seek to maximize its immediate potential utility gain, that is, to pass the resource to the neighbour valuing it the most.

Our main result is a bound in $\mathcal{O}(\Delta^2)$, where Δ is the max degree of the graph. Note in particular that this result does not depend on the number of agents. This upper bound for the expected length holds for all blind strategies, and the natural "utility maximizing" strategy mentioned above. But the argument developed in the proof of the theorem may not hold for some very specific strategies. For instance, one may imagine a heuristic consisting instead in taking the *min* among potential buyers.

4 Persuading Others

Properties of systems where agents exchange arguments to justify their opinions.

In previous chapters we assumed that preferences over alternatives (candidates, bundles of items) were given from the beginning, and would not change during the process. It is clear that in reality these preferences are based on various factors, one of them being the beliefs of the agents about the state of the world. These beliefs may vary during the decision-making process (for instance, an agent may initially believe two items to be complementary, but realize this is not the case once it acquires them). Or, alternatively, agents may discuss prior to decision-making and exchange relevant information (for instance, agents may discuss the respective strengths and weaknesses of candidates before proceeding to a vote). In such processes, agents will exchange expressive messages carrying arguments, that is, pieces of evidence that support their opinions. Sometimes the objective is to settle a debate, as in the case of persuasion we shall discuss later. But even if the ultimate goal of agents is not to reach a consensual outcome, the process leads to more informed decisions being made, thanks to the exchange of justifications which occurred. For instance, it is argued that discussing prior to voting may lead to more "single-peaked" shaped preferences of agents [LD03]. As a consequence, models of deliberation can be seen as complementary to those based on social choice theoretic approaches. I am a strong believer of this view, which makes a natural link between the various approaches discussed in this document.

Before we get further in the details, it is important to distinguish two different contexts which may lead to agents exhibiting different opinions on the same issues.

- in the case of *incomplete knowledge* about the state of the world, agents may hold different beliefs and/or preferences because they only have access to partial information. In theory, if it was not for the communication constraints, agents could exchange all (relevant) pieces of informations and agree upon the consensual outcome.

- in the case of *conflicting opinions*, agents may have intrinsically different beliefs and/or preferences. There is no necessarily such a thing as a "consensual" view *per se*, and agents may wish to convince other agents to adopt their own view: in this case it is they are typically engaged in a persuasion dialogue [Pra05].

A very distinctive feature of the approaches discussed in this chapter is the *non-monotonicity* of the reasoning process of the agents involved in the interaction, which means in the context of this chapter that we shall study situations where agents may change their beliefs on the basis of further information being gathered (this may subsequently trigger a modification of their preferences as mentioned above, but these aspects are not considered here—so we do not deal with *argument-based negotiation* for instance, where preferences or high-level goals of agents are typically affected by the various arguments put forward during a negotiation, see *e.g.* [RRJ+03, DMA08, HDM10]).

There are many approaches to non-monotonic reasoning in AI, but our contributions build on agents than can perform abduction and argumentation.

4.1 Background: Abduction and Argumentation

The works presented in this section assume that agents have beliefs that can change over time. More precisely, these beliefs are non-monotonic in the sense that some conclusion drawn from a knowledge base are not necessarily maintained when the knowledge base grows [BNT07].

4.1.1 Abduction

Abduction was defined by Peirce as the "probational adoption of a hypothesis" (quoted after [KKT93]) explaining some observations. There are different approaches to abduction but we shall only be concerned here with *logic-based abduction*.

Let $\mathcal{T}$ be the *background theory*, and O be an *observation*. At this point we do not commit to any specific language, so these are just respectively a set of sentences and a sentence. Given $\mathcal{T}$ and O, the abduction problem is first to find a set H of literals such that:

(i) $\mathcal{T} \cup H \models O$ (accountability),

(ii) $\mathcal{T} \cup H \not\models \bot$ (consistency)

However, this alone is not restrictive enough, as we typically want to constrain hypotheses to belong to a certain class of sentences that are candidate explanations (and not other effects, for instance). Thus, we define *Abd* as the set of *abducibles* (typically, a set of literals), which are candidate assumptions to be added to $\mathcal{T}$ for explaining O.

(iii) H is a set of sentences from *Abd* (bias).

An hypothesis H meeting conditions (i—iii) is also called an *explanation* of O (with respects to $\mathcal{T}$ and *Abd*). There are typically many candidates hypotheses for a given (set of) observations, which requires additional criteria to be able to discriminate and select some (or one, as we assume here) preferred hypothese(s). Some are non-controversial, as for instance the requirement of minimality which asks that we should not select an explanation subsumed by another explanation. On top of that, and depending on the domain considered, some additional criteria may be prove more adequate than others [Poo89]: for instance in a medical diagnosis problem we usually favour hypotheses exhibiting minimal cardinality, but this is not necessarily true in other contexts.

In this document we rely on two different systems for computing abductive explanations. The first one is the Theorist system developed by Poole [Poo89]. Equipped with this reasoning engine, agents can perform within the same framework abduction and prediction.

The second one is SOLAR [NIIR10], a deductive reasoning system based on SOL-resolution, which makes use of various sophisticated pruning techniques in order to find (minimal) consequences wrt. a given bias (or *production field*)[1]. Furthermore, SOLAR deals with (first-order) full clausal theories (so, no restricted to be Horn clauses).

4.1.2 Argumentation

Argumentation is an approach to non-monotonic reasoning which have received a great deal of attention over the recent years. In particular, the work of Dung [Dun95] was influential because it sets up an abstract framework allowing a great flexibility of analysis. The contribution discussed in this document builds on this trend, and we briefly introduce the key notions that we use.

Definition 4.1.1 *An argumentation system is a pair $\langle Arg, Att\rangle$, where Arg is a (finite) set of arguments, and Att is the attack relation, a binary relation over Arg.*

The approach is abstract in the sense that it does not commit to any specific internal content for argument, but only focuses on attack relations that exist among these arguments[2]. The objective is then to define what set(s) of arguments should be "collectively accepted". Dung [Dun95] proposes several possible semantics, which rely on a notion of "internal stability" called *conflict-freeness* (a set should not be self-contradictory and include arguments that attack each other), and various notions of "external stability" defining how the set should interact with arguments outside this set.

Definition 4.1.2 (Collective Defense) *Let $AS = \langle Arg, Att\rangle$ be an argumentation system. A set of arguments S collectively defends an argument a if, for all $b \in Arg$ such that $(b, a) \in Att$, it is the case that there exists $c \in S$ such that $(c, b) \in Att$.*

In this document, we shall just make use of the *grounded semantics*, a semantics which has the property to be unique and to always exist (although it may be empty). It turns out that the grounded semantics can be defined has being the least fixed-point of the function returning for a set of arguments, the set of arguments that it defends collectively.

Definition 4.1.3 (Grounded Semantics) *Let $AS = \langle Arg, Att\rangle$ be an argumentation system, and let $S \subseteq Arg$. S is grounded extension of AS iff S is the least fixed-point of the function $\mathcal{F} : 2^{Arg} \rightarrow 2^{Arg}$ with $\mathcal{F}(S) = \{a | S$ collectively defends $a\}$.*

4.1.3 Distributed Approaches to Abduction and Argumentation

Now, as in the rest of this document, we shall be concerned with distributed approaches to abduction and argumentation. Several works have investigated similar issues, in particular in recent years. As mentioned in the introduction already, it is crucial to make a clear distinction

[1] Consequence finding, unlike theorem proving, seeks to deduce 'interesting' theorems from a given set of axioms (see [Mar00] for a survey of dedicated algorithms). It can be used for abductive reasoning by using the principle of *inverse entailment*, which makes use of the fact the accountability condition can be rewritten as $\mathcal{T} \cup \{\neg O\} \models \neg H$, and that the consistency condition can be rewritten as $\mathcal{T} \not\models \neg H$, so we are looking for 'new' theorems obtained when adding the negation of observations to the background theory.

[2] We should mention that this is certainly not the only approach to argumentation, see for instance [BH08] or [RS09] for a recent overview of the field.

between approaches which study to what extent a given procedure can be distributed (even maybe seen as a "virtual" interaction between agents), and those which investigate instead the properties of systems populated by non-cooperative agents endowed with abductive and argumentative properties.

This notice is especially important in argumentation where the use of somewhat ambiguous terms may be confusing for the reader. For instance the various proposals of dialectical proof-procedures for argumentation, interpreted as *argument games* [PS97] between a proponent and an opponent, clearly fall in the first category [Pra05]. In this case, we certainly want the issue of the debate to be pre-determined from the initial situation for the proof-theory to be sound and complete. It differs from the case where agents, equipped with argumentation systems, are allowed by an appropriate protocol [Bre01, PSJ98, AMP00, APM00] to exchange arguments, in a way that is strategical and which may lead to different outcomes to be reached [PWA03]. In this case the protocol needs to mediate the interaction in a way that is both efficient and "fair" [Lou02], as advocated already by the "computational dialectics" approach [Gor96]. In the context of persuasion that we shall discuss here, fair may mean that agents indeed have a chance to convince other agents, which mean in turn that the way they play may make a difference regarding the outcome. These strategical issues of distributed argumentation have recently been put forward in a series of papers [RL09, CP10] where the authors take a mechanism design perspective on the multiagent argumentation problem. The work described in Section 4.3 follows this trend and proposes a multiparty (as opposed to bilateral) protocol for persuasion.

In contrast, the work on distributed abduction presented in the next section is the only one which departs from the assumption of non-cooperativeness made throughout the document. In this case indeed, agents have no individual interest in trying to impose their opinion: they just want to cooperatively contribute to the collective explanations being build. For instance, they would not hide an observation even though they believe or know that this information will later be damaging for the hypothesis they defend, nor would they specifically select a given counter-example because they think it favours the opinion they currently hold. This is in line with several approaches developed in recent years, that have put forward the frameworks of ALIAS [CLM+03] and DARE [MRBC08]. A noticeable difference is that our work puts a special focus on communication constraints that are imposed by the network topology, in the sense that the system is not necessarily fully connected, and that agents are confined to use bilateral interactions. The framework of DeCA [ACG+06] considers similar constraints but deals with propositional abduction.

4.2 Distributed abduction: Agents with Incomplete Knowledge

Take n agents of a system, making observations in a world globally consistent (meaning that with full knowledge of the observations of all agents, it would be possible to build a satisfying hypothesis). Suppose now that each agent holds an incomplete view: distributed on a network, they may only locally perceive their environment, and build a hypothesis of the current state of the system that fits their local view. In order to refine their hypothesis, they will communicate with other agents that may be reached in the network. The question that we address here is the following: how will such a system evolve for a given protocol? Will it converge to a state globally satisfying?

More formally, a *multiagent abductive system* is defined as a tupe $\langle \mathcal{N}, G_t, A \rangle$, where:

- $\mathcal{N}$ is a set of agents and each agent is a pair $\langle T_i, O_i \rangle$ such that T_i is the *individual theory* of agent i, and O_i is his set of *observations*. The agent can build hypotheses that, together with T_i can explain the observation set O_i.
- Π_G is a sequence of communicational constraints graphs G_t. More precisely, at each time step t, the (undirected) graph $\langle \mathcal{N}, E_t \rangle$ specifies which agents can communicate with each others.
- *Abd* is a common set of abducibles
- $\succ$ is a common preference relation, a linear order over the possible hypotheses that may be produced.

Theories and observations are considered as certain knowledge, and following this the whole system is consistent, that is $\cup_{i \in \mathcal{N}}(T_i \cup O_i) \not\models \bot$. Another important assumption made here is that agents are always able to single out a single preferred hypothesis h_i on the basis of the information available to them at a given time. As mentioned above, we seek to design protocols that can ensure convergence to states that are satisfying at the global level. But we need to define what these states are.

4.2.1 Group-accountability and group-consistency

In the context of hypothetical reasoning, its is intuitive to measure to what extent the different hypotheses that the agents come up with are *explainable* with respect to the whole system. Abstractly, this notion of explainability may just be conceived as a binary relation linking a set of observations to an hypothesis. As mentioned before, explainability in the context of abductive reasoning considered here consists of two requirements: the hypothesis should explain all the observations of the agent, and it should be consistent with the background knowledge of that agent. It is easy to extend these definitions to a group of agents. In this case, we say that:

- h is *group-consistent* with G if it is consistent with the union of all the individual theories, that is, $\cup_{i \in \mathcal{G}} T_i \cup h \not\models \bot$.
- h is *group-accountable* with G if (together with the theories of the group) it can account for all the observations made in the group, that is, $\cup_{i \in G} T_i \cup h \not\models \cup_{i \in G} O_i$.

By extension, when h is both group-consistent and group-accountable for G we say that it is an *group-explanation* for G.

Now we need to characterize the state of the system. Remember that each agent builds his own hypothesis, how satisfying is the set of hypothesis obtained as a whole? This gives rise to different levels, depending on how "similar" are the hypotheses, and on the scale used to appreciate group-explanations. We say that a group of agents G is:

- *mutually explainable* when all agents i hold an individual hypothesis h_i which is a group-explanation for G.
- *homogeneous* when the group is mutually explainable, and all hypotheses belong to the same equivalence class (that is, when they explain exactly the same sets of observations).

Finally, at the level of the system, we may only require all *neighbours*, or more demandingly all agents to exhibit the desired property. We say that a system is:

- *locally* mutually explainable (resp. homogeneous) when all pairs of agents $(i, j) \in E_t$ are mutually explainable (resp. homogeneous).
- *globally* mutually explainable (resp. homogeneous) when the system itself is mutually explainable (resp. homogenous).

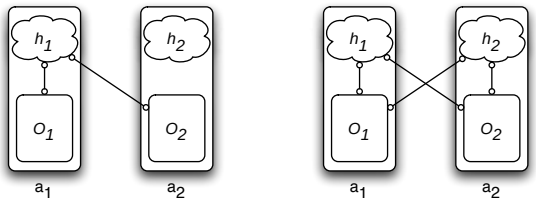


Figure 4.1: Group-explainability (left) and mutual peer explainability (right) of agents a_1 and a_2.

Figure 4.1 illustrates these notions (a link between h and O means that $Expl(h, O)$ holds.): on the left-hand side, the hypothesis h_1 is group-explainable with the group $\{a_1, a_2\}$. However this group of agents itself is not mutually explainable since h_2 is not a group-explanation for it. On the right-hand side, the group is mutually explainable. Note that mutual explainability does not necessarily imply that each agent holds the *same* hypothesis. It is perhaps also useful to observe that the notion of local mutual explainability is not transitive, which means that it may not easily level up to global mutual explainability.

Example 2 *We construct a simple counter-example, where the following theory T is shared by all agents.*

$$\left\{ \begin{array}{lll} h_1 \rightarrow o_1, & h_2 \rightarrow o_1, & h_3 \rightarrow o_2, \\ h_1 \rightarrow o_2, & h_2 \rightarrow o_2, & h_3 \rightarrow o_3, \\ h_1 \rightarrow \neg o_3, & h_2 \rightarrow o_3 & h_3 \rightarrow \neg o_1 \end{array} \right\}$$

Thus, h_2 can explain all observations, h_1 is consistent with the observation set $\{o_1, o_2\}$, and h_3 is consistent with observation set $\{o_2, o_3\}$. Now the system is in the state $\{\langle h_1, \{o_1\}\rangle, \langle h_2, \{o_2\}\rangle, \langle h_3, \{o_3\}\rangle$, and $E = \{(1, 2), (2, 3)\}$. In other words, agent 1 cannot communicate with agent 3. We see that the system is locally mutually explainable, since each pair of neighbours taken separately is mutually explainable. However the property is not transitive and one can verify that $(1, 3)$ is not mutually explainable, since their hypotheses are not consistent with the observations made by the other. Hence the system is not globally mutually explainable.

The challenge of this work is precisely to design protocols that give some guarantee that states where the system is globally mutually explainable can be attained.

4.2.2 Compositionality of explanations

Just like in chapter 3 where the domain of preferences used by agents affected the complexity of the protocol required to reach satisfying states, in the context of this chapter one can make some assumptions on the explainability relation. Essentially, the more "modular", the more global properties can be ensured by means of local and independent computations.

In general (at least if the underlying entailment relation is classical as we assume here) the following properties are satisfied:

- if h is group-consistent for G, then it is the case that for any $G' \supseteq G$, h is group-consistent for G'. Another way to put it is to say that non-consistency is monotonic: when an agent finds a counter-example for a given hypothesis, then this hypothesis can be ruled out for good.
- if h is group-accountable for both G and G', then it is group-accountable for $G \cup G'$. This means that group-accountability can be checked locally, and we say that accountability is additive [Fla96].

We may work in domains where a much stronger condition is fullfilled, namely where the explainability relation as a whole is *compositional*, meaning that:

$$Expl(h, O) \text{ and } Expl(h, O') \Leftrightarrow Expl(h, O \cup O')$$

In these domains, not only the accountability is additive but the consistency as well. For instance, it rules out domains where one agent cannot discard an hypothesis on the basis of two observations taken separately, but where he could if he received both observations. These domains are certainly much easier to deal with, since it basically means that it is possible to consider independently each observation.

4.2.3 Protocols for hypothesis refinement

The general idea of these protocols is based on a learner/critic approach.They allow agents to propose hypotheses, but also to contradict the hypothesis of others by providing observations that may serve as counter-examples. As we restrict ourself to bilateral interactions, there are actually two layers of protocols to consider. At the local level, we design protocols that regulate bilateral interactions between agents. At the global level, we design protocols that regulate how the local protocols are triggered. In a series of paper, we studied this problem in a number of different contexts. We started in compositional domains, first with systems where the communication links are static (for all t, $E_t = E_{t+1}$), then with systems where the links change over time [BHMP07, BSM09]. More recently, we investigated non-compositional domains [BIM10a]. This line of work was initiated with the co-supervision of the PhD thesis of Gauvain Bourgne, and is currently being pursued as Gauvain is now a postdoc researcher in the group of Katsumi Inoue (NII, Tokyo). We now briefly skim through the key results obtained throughout this collaboration.

4.2.4 Compositional domains

To ensure that domains are compositional, a requirement is that all agents share the same background theory, *i.e.* there is no pair of agents (i, j) such $T_i \neq T_j$. (This is not a sufficient condition though. More generally the exact characterization of compositional domains is

the subject of current research). In compositional domains, the basic building block is a protocol for hypothesis exchange which allows each agent, in turn, to propose and contradict hypothesis. This simple basic protocol is depicted in Figure 4.2.

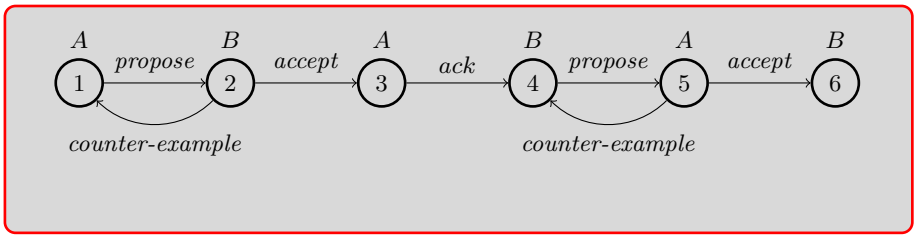


Figure 4.2: A protocol for hypothesis exchange

One can see in particular that in state 4, the roles are swapped (B becomes the learner). It is very instructive to provide a concrete instantiation of a strategy that may be used together with such a protocol, to see what kind of informations are exchanged. Upon receiving a hypothesis h_1 ($propose(h_1)$) from a_1, agent a_2 is in state 2 and has the following possible replies:

- if $\exists o_2 \in O_2$ s.t. $h_1 \models \neg o_2$ or $\exists o_2 \in O_2$ s.t. $h_1 \not\models o_2$, then agent a_2 has an observation which is not explainable with hypothesis h_1: it will then communicate this counter example by replying $counter\text{-}example(o_2)$). We are back in state 1 of the protocol. Agent a_1 will then update its hypothesis with this new observation and *propose* h'_1.
- Otherwise, we have $Expl(h_1, O_2)$. In this case, agent a_2 send $accept(h_1)$ and a_1 acknowledges this with *ok*. We are now in state 4. The role of a_1 and a_2 are swapped. Upon receiving h_2, agent a_1 will reply with *counter-example* or *accept* using the same criterion that a_2 used in state 2, until we have $Expl(h_2, O_1)$.

Note that there are two kinds of counter-examples that can be provided by the agent playing the role of the critic: observations that show the hypothesis to be inconsistent, and observations that the hypothesis fails to account for. In compositional domains, consistency check and accountability check can be performed individually by each agent. It is simple to show as we do in [BSM09] that this bilateral protocol and the associated strategy guarantee a mutually explainable (in fact, homogeneous) state to be reached.

Convergence in static settings. In collaboration with Gauvain Bourgne and Amal El-Fallah Seghrouchni, we first considered the case of static communication links in the system. In [BSM09] we discuss a large variety of protocols, used either to regulate local bilateral interactions, or global interactions. An interesting case is provided by the global protocols required to deal with non fully connected systems: in this case, it is necessary to rely on a principle of propagation since a single cannot reach all the other agents in the system. The proposed solution amounts to build on-the-fly a spanning tree: indeed upon receiving a new observation, an agent will consider his neighbours as children, will apply a local protocol with the first of them and ask him, in turn, to propagate the information to its own (not already linked in the tree) neighbours, and so on. When a counter-example is provided, it is propagated back to the root, and the process starts again with a new hypothesis. This

global protocol provably converges (under some assumption regarding agents' behaviour) to a globally mutually explainable state of the system.

Convergence in dynamic settings. In dynamics settings, convergence can still be ensured, but the techniques used in the static can no longer be employed. Instead, hypotheses must be propagated despite the lack of knowledge regarding the communication links that will exist in the next time step (in a way that is somewhat reminiscent of a rumour). In joint work with Gauvain Bourgne, Suzanne Pinson, and Gael Hette [BHMP07], we investigated some sufficient conditions allowing to guarantee convergence despite unpredictable communication links. The first requirement is that the system is *temporally connected*. Intuitively, it means that there is (asymptotically) always a way, for any agent, to find a temporal path of communication so as to reach any other agent. This means that at any time t, the graph $\langle \mathcal{N}, \cup_{t' \geq t} E'_t \rangle$ must be a connected graph. (Note that it does not guarantee at all that a given communication link between two agents will appear at some point in the future). Then a sufficient condition, for each pair of agents temporally connected to be eventually group explainable, would be that each local interaction guarantees a *transitive* explainability relation to hold (and it turns out that only homogeneity can guarantee this). Finally agents must be willing to trigger bilateral interactions that are *relevant* (which means in particular that their strategies must be designed in such a way that those interactions between already homogenous agents are recognized and avoided), and the global protocol must be *non-blocking* (*i.e.* not designed in such a way that it precisely prevents those relevant interactions to take place).

Experiments in non-temporally connected systems. There is no doubt that the assumption of temporal connectivity of the system is a strong one. To take a step further into the direction of a realistic application, we did set up an experimental case-study simulating a crisis scenario where a number of agents try to escape a building in fire that we first described in [BMP06]. In this application, agents where equipped with a reasoning engine provided by the Theorist system of Poole [Poo89]. The application is untypical is that it mixes (abductive) reasoning (finding out where the origin of the fire might be), prediction (based on this hypothesis, knowing where the fire will propagate in the future), and acting (based on these predictions, chosing the best route to exit). The application does not guarantee temporal connectivity because agents may "die" when caught by the fire. One of the main question was whether the hypothesis refinement protocol would perform well when compared to simpler observation exchange protocols. These protocols were tested with a number of instances, varying in particular the difficulty for agents to exit the building. The main conclusion drawn is that seeking mutual explainability of hypothesis is only important for maps that include potentially critical situations, that is, states where failing to locate *exactly* the fire origin may lead to unrecoverable bad actions.

4.2.5 Non-compositional domains

In recent joint work with Gauvain Bourgne and Katsumi Inoue, we started to investigate the case of agents equipped with full clausal theories, and this time not sharing the same background theory. This time the domain is not compositional since (*i*) consistency is not additive (so coherence checking must be performed collaboratively by agents), and (*ii*) ac-

countability is not incremental (so "false" counter-examples for accountability are possible, and justifications must be attached to counter-examples).

In [BIM10a] we design a sophisticated bilateral protocol for distributed abduction. The protocol is also following a learner/critic approach, and incrementally builds a *validity context* attached to each discussed hypothesis. The interaction starts with a learner agent (say a_0) putting forward a candidate hypothesis h_0 accompanied by its validity context to the critic agent (say a_1). The protocol is then composed of four phases:

(i) a *consistency check* phase, which lets the critic agent check consistency wrt. its own theory and, if no inconsistency is detected, augment the context with new consequences derived from $h_0 \cup T_0 \cup T_1$, and send it back to the learner who in turn add newly derived consequences to the context, and so on. This loops until no new consequences are found or an inconsistency is detected.

(ii) an *accountability check* phase, where the critic agent checks whether all its observations are explained by $h_0 \cup T_1$. If an observation o not accounted for is found, it is sent back to the learner. Now if $h_0 \cup T_0$ does not explain it neither, o is indeed a counter-example. But if $h_0 \cup T_0$ *do* account for o, the learner will provide an "argument" by providing the part of the hypothesis which is useful to explain o and the critic adds the clause $\neg \hat{h} \vee o$ to his theory.

(iii) an *acceptability check*, during which other candidate hypotheses are explored if necessary (in particular if the hypothesis is partial, as defined below).

(iv) an *acceptation phase*, during which the best hypothesis is eventually selected.

Because an agent may not be able to produce an explanation fully composed of abducibles (remember that each agent may have part of the background theory at his disposal), the protocol allows *partial hypotheses* to be exchanged, that is, hypotheses completed by non-abducible (*e.g.* other literals of the language).

Example 3 *[BIM10a] There are two agents, the language of abducibles is* $Abd = \{g(X), i(X)\}$. *Moreover,* $O_0 = \{\}$ *and* $O_1 = \{b(c_2)\}$. *Initially,* a_0 *has hypothesis* $h_0^0 = \{\}$ *and* a_1 *has hypothesis* $h_1^0 = b(c_2)$ *as it has no clause containing* $b(X)$.

T_0	T_1
$b(X) \vee \neg g(X)$	$\neg e(X) \vee \neg c(X)$
$\neg a(c_1)$	$h(c_1)$
$b(X) \vee \neg k(X)$	$a(X) \vee c(X) \vee \neg d(Y, X)$
$e(X) \vee \neg h(X)$	$k(X) \vee \neg i(X)$
$f(X, Y) \vee \neg h(Y) \vee \neg g(X)$	$d(X, Y) \vee \neg f(X, Y)$

If agent a_0 *starts the interaction, it first proposes its empty hypothesis, which goes directly through the consistency check of* a_1 *since it has no new consequences. Then in the accountability step,* a_1 *sends* $b(c_2)$ *as a counter example, and* a_0 *has to compute a new hypothesis. It gets* $h_0^1 = g(c_2)$*, which will be proved inconsistent in the consistency check, resulting in the computation of a new hypothesis,* $h_0^2 = k(c_2)$*, which succeeds the consistency check. Then, during accountability check,* a_1 *considers that* $b(c_2)$ *is not explained by* $k(c_2)$ *and sends it as a counter-example. Since* T_0 *can explain this observation with* $k(c_2)$*, it returns* $(b(c_2) \vee \neg k(c_2))$ *as an argument (note that the rule is then instantiated). This rule is added to the theory of*

a_1, and as there are no other observations, a_1 moves to the acceptability check. Since $k(c_2)$ is not an abducible, it will not be accepted before exploring other hypothesis. h_1^0 is not preferred over h_0^2, but since a_1 got a new rule, it can use it to compute new hypothesis. As a result it gets $h_1^1 = i(c_2)$, which respects the bias condition. Thus h_0^2 is temporarily denied and ruled out, before a_1 proposes $i(c_2)$ with the newly computed context $\{i(c_2), k(c_2)\}$. The context is confirmed, but during the accountability check, a_0 sends $b(c_2)$ as a counter-example. a_1 thus argues $(b(c_2) \vee \neg i(c_2))$ (grounding and combining its clauses), which is added to the theory of a_0. Then, during acceptability step, a_0 cannot compute a better hypothesis than h_1^1. This hypothesis is composed of abducibles, so there is no need to explore further, and it is accepted, ending the exchange.

To ensure termination of these local interactions, we have to make sure that the number of possible hypotheses and context that can be derived from the theory and observations is finite. As in the domains discussed before, this bilateral protocol is coupled with a global protocol that we do not detail again here, and with adequate strategies, which allows to study their properties.

Theorem 4.2.1 (Soundness [BIM10a]) *The system is globally mutually explainable wrt. to any hypothesis accepted by all agents (with the same validity context) as a consequence of the protocol.*

Provided that termination of local interactions is guaranteed as discussed before, that the system is temporally connected, and by the soundness result just given, the multiagent abductive system will converge to a state where every group of connected agents is mutually explainable (in fact, homogeneous and sharing the same hypothesis). Now, this does not rule out that the hypothesis obtained by this process is in fact a partial hypothesis (instead of one fully composed of abducibles). The next result states that this is not the case.

Theorem 4.2.2 ((Partial) Completeness [BIM10a]) *If there exists at least one global explanation (wrt. Abd) then the system converges to a state where all agents share such an explanation.*

In [BIM10b] further refinements are discussed to speed-up the computation (the many calls to the consequence finding system are very demanding) and the communication overhead of the protocol. These refinements are based on various heuristics used to avoid redundant computation and select carefully the more promising interactions: for instance, assuming full knowledge of the different languages used by agents, it is possible to select wisely bilateral interactions.

4.3 Multiparty argumentation: Agents with Conflicting Opinions

We now turn our attention to the situation where agents hold different conflicting views of the world. We shall assume to start with that each agent is modeled as an argumentation system in Dung's style, and denote by $\mathcal{E}(AS_i)$ the arguments accepted under the grounded semantics for agent i. Now each agent may have different opinions as to whether some attacks hold between arguments. Suppose indeed we are in the presence of two mutually conflicting arguments, say α and β. The agents may have different opinions regarding the respective credibility of

a given source, hence in one argument system the attack would hold from α to β, while the symmetric relation would hold for the other agent. More generally, the situation may occur as the result of agents being equipped with preference-based (or value-based) argument systems [Bre01, BC02] sharing the same arguments but diverging when it comes to the preferential value attached to arguments. The aim is here is to design a multiparty (as opposed to simply bilateral) protocol that regulates satisfyingly persuasion dialogues among many agents. We assume that agents are *focused* [RT10], that is, they concentrate their attention on a specific (same for all) argument. This argument is referred to as the *issue* d of the debate [Pra05]. Unsurprisingly, agents want to see the acceptability status (under the grounded semantics) of the issue coincide in the debate and in their individual system. It is therefore useful to see the debate as opposing two groups of agents. For that purpose we introduce two sets of agents: $CON = \{a_i \in N | d \notin \mathcal{E}(AS_i)\}$ and $PRO = \{a_i \in N | d \in \mathcal{E}(AS_i)\}$. This is ongoing work with Elise Bonzon, Université Paris-Descartes.

4.3.1 Outcome of Merged Argumentation System

The mere definition of what outcome the designer should aim for at the level of the society is unclear. This is so because aggregating a set of conflicting argumentation systems is already challenging, and may give rise to a number of different solutions. We rely on the work of Coste-Marquis et al. [CMDK+07] who investigated this problem. In the specific case we discuss here, it turns out that a meaningful way to merge is to take the *majority argumentation system* where attacks supported by a majority of agents are kept (this corresponds to minimizing the sum of the edit distances between the AS_i and the merged system [CMDK+07]). Assuming on top of that ties are broken in favour of the absence of an attack allows to ensure to existence of a single such merged argumentation system, that we denote MAS_N.

Definition 4.3.1 *Let $AS_1 \dots AS_n$ be n argumentation systems. The* majority argumentation system *is $\langle A, M \rangle$ where $M \subseteq A \times A$ and xMy when $|\{i \in N | xR_iy\}| > |\{i \in N | x \not R_i y\}|$.*

From now on we refer to $\mathcal{E}(MAS_N)$ as the *merged outcome* for a set of agents N.

4.3.2 A Simple Relevance-based Protocol for Multiparty Persuasion

We follow [Pra05] and use *direct relevance* as a myopic criteria to restrict the range of acceptable moves (just like individual rationality in the context of resource allocation discussed in Chapter 3). What this means is that a move is acceptable if it immediately modifies the current status of the issue under discussion. Another important feature is that the protocol allows only one argument to be advanced at the same time. The permitted moves are as follows: either positive assertions of attacks xRy, or contradiction of (already introduced) attacks. When a (relevant) move is played on the gameboard, the following update operation takes place:

1. after an assertion xRy
   - if $xR_\alpha y \in A^t(GB)$ then $\alpha \leftarrow \alpha + 1$
   - if $xR_\alpha y \notin A^t(GB)$ then the edge is created with $\alpha \leftarrow 1$
   - otherwise (either x or y is not present), then the node of the new argument is created and the edge is created with $\alpha := 1$

2. after a contradiction $x \not R y$ iff $x, y \in A^t(GB)$, then $\alpha \leftarrow \alpha - 1$

All the ingredients are in place to describe the simple protocol which will regulate the multiparty debate.

---

1. Agents report their individual view on the issue to the central authority, which then assign (privately) each agent to PRO or CON.
2. The first round starts with the issue on the gameboard and the turn given to CON.
3. Until a group of agents cannot move:
   a) let agents independently propose moves to the central authority;
   b) the central authority picks the first (or at random) relevant move from the group of agents whose turn is active, update the gameboard, and passes the turn to the other group

---

Table 4.1: A multiparty argumentation protocol

The type of questions that we are now interested in are the ones encountered many times already in this document: for instance, can we guarantee that the system will converge to the merged outcome (or at least that such a state is reachable)? Is there room for strategy in a protocol defined like this (*i.e.* is the outcome pre-determined from the initial situation)? We just give a glimpse of the results obtained so far, to illustrate these ideas.

4.3.3 Lack of Reachability of the Merged Outcome

The example of Figure 4.3 shows that the outcome[3] of the merged argumentation system may not be reachable, if agents follow the simple protocol described in Table 4.1.

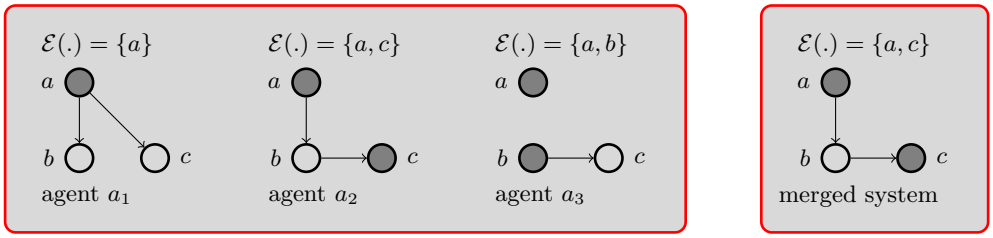


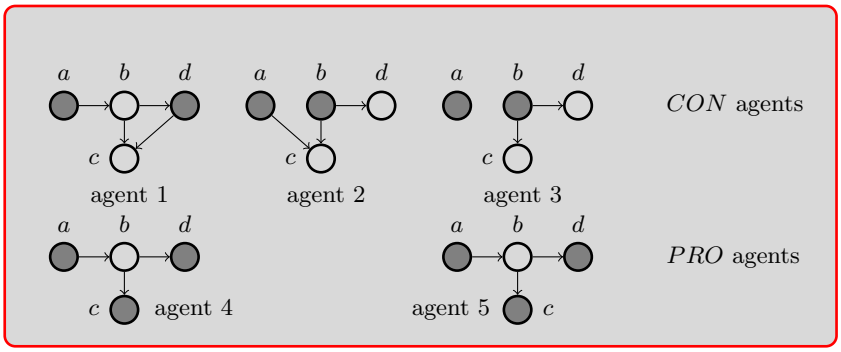
Figure 4.3: Three agents with their argumentation systems, and the merged system

This means that from the initial state consisting of the empty gameboard, no sequence of relevant moves performed according to the protocol can lead to a state where c would be *in*. Intuitively, one sees that a key element here is that the merged outcome requires that a_1 plays the move aRb to be established. However this is never in a_1 interest to do so (as a_1 is against the issue c).

[3]Observe that c is *in* in the merged argumentation system, but that a majority of agents (a_1 and a_3) have c as *out* in their extension.

4.3.4 The Room for Strategy

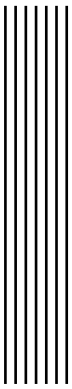
The next question is whether our protocol actually leaves some degree of freedom to agents (that is, whether strategy may make a difference), or whether the issue of the debate is pre-determined from the initial situation. The answer is not be obvious at first sight, but the protocol is *not* pre-determined. To illustrate this we exhibit an example where both sides may win the debate (the issue is c).



Example 4 *Five agents have the following argumentation systems: We can see that the issue c is a* possible outcome *for the agents in CON: CON can attack c with b. Then, the only possible move for PRO is to defend d with aRb. However, CON can remove this attack, and PRO has no other move. But more surprisingly c is also a* possible outcome *for PRO: CON can start by adding (d, c), which is playable by a_1. Then, a_5 will defend with (b, d), and a_1 counter-attack with (a, b). If the next move of PRO is to remove the attack (d, c), then CON has no other move left: it cannot add the attack (b, c), as it is defended by a; and it cannot remove the edge (a, b) as it does not drop c from the extension. In this case, (a, b) is a "switch" and the merged outcome is then (only) reachable.*

In ongoing work we seek an exact characterization of debates that are indeed "open", and rely for this on an analysis of what team of agents gets the power to establish or contradict attack relations (giving rise to an *argument-control graph*). Several refinements are possible, for instance a delicate problem is how agents are supposed to react upon receiving an argument they were not aware of before. In this case, different models could prove worth investigating, for instance those accounting for the notion of *influence* among agents.

The issues sketched in this last section give an idea of the possible developments I wish to have the opportunity to investigate during the co-supervision of the PhD thesis of Dionysios Kontarinis.



Bibliography

[ACG+06] P. Adjiman, P. Chatalic, F. Goasdoué, M.-C. Rousset, and L. Simon. Distributed reasoning in a peer-to-peer setting: Application to the semantic web. *Journal of Artificial Intelligence Research (JAIR)*, 25:269–314, 2006. 48

[ADG91] A. Alkan, G. Demange, and D. Gale. Fair allocation of indivisible goods and criteria of justice. *Econometrica*, 59(4):1023–1039, 1991. 34

[AE10] S. Airiau and U. Endriss. Multiagent resource allocation with sharable items: Simple protocols and nash equilibria. In *Proceedings of the 9th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2010)*, May 2010. 30

[AMP00] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS00)*, pages 31–38, Boston, USA, July 2000. IEEE Press. 48

[APM00] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the European Conference on Artificial Intelligence (ECAI-2000)*, pages 338–342, Berlin, Germany, August 2000. IOS Press. 48

[Arr51] K. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 1951. revised edition 1963. 17

[BBD+04] C. Boutilier, R. I. Brafman, C. Domshlak, H. H. Hoos, and D. Poole. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *J. Artif. Intell. Res. (JAIR)*, 21:135–191, 2004. 32

[BBF10] Y. Bachrach, N. Betzler, and P. Faliszewski. Probabilistic possible-winner determination. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 2010. To appear. 20

[BC02] T. Bench-Capon. Value-based argumentation frameworks. In *NMR'02*, pages 443–454, 2002. 56

[BEL09] S. Bouveret, U. Endriss, and J. Lang. Conditional importance networks: A graphical language for representing ordinal, monotonic preferences over sets of goods. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-2009)*, pages 67–72, July 2009. 31

[Bev98] C. Beviá. Fair allocation in a general model with indivisible goods. *Review of Economic Design*, 3(3):195–213, 1998. 34

[BH08] P. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008. 47

[BHMP07] G. Bourgne, G. Hette, N. Maudet, and S. Pinson. Hypotheses refinement under topological communication constraints. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'07)*, pages 994–1001. ACM Press, May 2007. 51, 53

[BHN09] N. Betzler, S. Hemmann, and R. Niedermeier. A multivariate complexity analysis of determining possible winners given incomplete votes. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, pages 53–58, 2009. 20

[BIM10a] G. Bourgne, K. Inoue, and N. Maudet. Abduction of distributed theories through local interactions. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)*, August 2010. 51, 54, 55

[BIM10b] G. Bourgne, K. Inoue, and N. Maudet. Towards efficient multi-agent abduction protocols. In *Third international Workshop on LAnguages, methodologies and Development tools for multi-agent systemS (LADS-2010)*, 2010. 55

[BL92] S. Berg and D. Lepelley. Note sur le calcul de la probabilité des paradoxes de vote. *Mathématiques et Sciences Humaines*, 120:33–48, 1992. 21

[BL08] S. Bouveret and J. Lang. Efficiency and envy-freeness in fair division of indivisible goods: Logical representation and complexity. *Journal of Artificial Intelligence Research*, 32:525–564, 2008. 34

[BMP06] G. Bourgne, N. Maudet, and S. Pinson. When agents communicate hypotheses in critical situations. In *Proceedings of the Fourth International Workshop on Declarative Agent Languages and Technologies (DALT–2006)*, May 2006. 53

[BNT07] G. Brewka, I. Niemela, and M. Truszczynski. Nonmonotonic reasoning. In F. van Harmelen V. Lifschitz, B. Porter, editor, *Handbook of Knowledge Representation*, pages 239–284. Elsevier, 2007. 46

[BR08] Y. Bachrach and J. S. Rosenschein. Distributed multiagent resource allocation in diminishing marginal return domains. In *Proc. 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, pages 1103–1110. IFAAMAS, 2008. 30

[Bre01] G. Brewka. Dynamic argument systems: A formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation*, 11(2):257–282, 2001. 48, 56

[BSM09] G. Bourgne, A. El Fallah Seghrouchni, and N. Maudet. Towards refinement of abductive or inductive hypotheses through propagation. *Journal of Applied Logic*, 7(3):289–306, 2009. Special Issue on Abduction and Induction in Artificial Intelligence. 51, 52

[BT96] S.J. Brams and A.D. Taylor. *Fair Division: From Cake-cutting to Dispute Resolution*. Cambridge University Press, 1996. 29

[BTT89a] J.J. Bartholdi, C.A. Tovey, and M.A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989. 17

[BTT89b] J.J. Bartholdi, C.A. Tovey, and M.A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(3):157–165, 1989. 17

[BTT92] J. Bartholdi, C. Tovey, and M. Trick. How hard is it to control an election? *Social Choice and Welfare*, 16(8-9):27–40, 1992. 25

[CDE+06] Y. Chevaleyre, P. E. Dunne, U. Endriss, J. Lang, M. Lemaître, N. Maudet, J. Padget, S. Phelps, J. A. Rodríguez-Aguilar, and P. Sousa. Issues in multiagent resource allocation. *Informatica*, 30:3–31, 2006. 30

[CDLS02] M. Cadoli, F. Donini, P. Liberatore, and M. Schaerf. Preprocessing of intractable problems. *Information and Computation*, 176(2):89–120, 2002. 21

[CEEM04] Y. Chevaleyre, U. Endriss, S. Estivie, and N. Maudet. Welfare engineering in practice: on the variety of multiagent resource allocation problems. In M. P. Gleizes, A. Omicini, and F. Zambonelli, editors, *Proceedings of the Fifth International Workshop Engineering Societies in the Agent World*, volume 3451 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 335–347, Toulouse, October 2004. Springer-Verlag. 29

[CEEM07] Y. Chevaleyre, U. Endriss, S. Estivie, and N. Maudet. Reaching envy-free states in distributed negotiation settings. In *Proc. of IJCAI-2007*, 2007. 40

[CEEM08] Y. Chevaleyre, U. Endriss, S. Estivie, and N. Maudet. Multiagent resource allocation in k-additive domains: Preference representation and complexity. *Annals of Operations Research*, 163(1):49–62, 2008. 36

[CELM05] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. Negotiating over small bundles of resources. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*, pages 296–302. ACM Press, July 2005. 36

[CELM07] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. A short introduction to computational social choice. In *Proceedings of the 33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM-2007)*, volume 4362 of *LNCS*, pages 51–69. Springer-Verlag, January 2007. 14, 17, 18

[CELM08] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. Preference handling in combinatorial domains: From ai to social choice. AI Magazine. Special Issue on Preferences., 2008. 8

[CEM05] Y. Chevaleyre, U. Endriss, and N. Maudet. On maximal classes of utility functions for efficient one-to-one negotiation. In *Proc. 19th International Joint Conference on Artificial Intelligence (IJCAI-2005)*, pages 941–946. Morgan Kaufmann Publishers, 2005. 38

[CEM06] Y. Chevaleyre, U. Endriss, and N. Maudet. Tractable negotiation in tree-structured domains. In P. Stone and G. Weiss, editors, *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006)*, pages 362–369. ACM Press, May 2006. 36

[CEM07] Y. Chevaleyre, U. Endriss, and N. Maudet. Allocating goods on a graph to eliminate envy. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI-2007)*, pages 700–705. AAAI Press, July 2007. 40, 41

[CEM08] Y. Chevaleyre, U. Endriss, and N. Maudet. Trajectories of goods in distributed allocation. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, pages 1111–1118. IFAAMAS, May 2008. 43

[CEM10a] Y. Chevaleyre, U. Endriss, and N. Maudet. Distributed fair allocation of indivisible goods. Working Paper, 2010. 34, 39, 40

[CEM10b] Y. Chevaleyre, U. Endriss, and N. Maudet. Simple negotiation schemes for agents with simple preferences: Sufficiency, necessity and maximality. *Journal of Autonomous Agents and Multiagent Systems*, 20(2):234–259, 2010. 38, 42

[CKKK09] I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou. The efficiency of fair division. In S. Leonardi, editor, *WINE*, volume 5929 of *Lecture Notes in Computer Science*, pages 475–482. Springer, 2009. 34

[CLM+03] A. Ciampolini, E. Lamma, P. Mello, F. Toni, and P. Torroni. Cooperation and competition in ALIAS: a logic framework for agents that negotiate. *Annals of Mathematics and Artificial Intelligence*, 37(1–2):65–91, 2003. 48

[CLM+10] Y. Chevaleyre, J. Lang, N. Maudet, J. Monnot, and Lirong Xia. New candidates welcome! possible winners with respect to the addition of new candidates. *Mathematical Social Science*, 2010. under revision. 25, 26

[CLMM10] Y. Chevaleyre, J. Lang, N. Maudet, and J. Monnot. Possible winners when new candidates are added: the case of scoring rules. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 2010. 25, 26

[CLMRA09] Y. Chevaleyre, J. Lang, N. Maudet, and G. Ravilly-Abadie. Compiling the votes of a subelectorate. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-2009)*, pages 97–102, July 2009. 21, 22

[CLS03] V. Conitzer, J. Lang, and T. Sandholm. How many candidates are required to make an election hard to manipulate? In *Proceedings of TARK-03*, pages 201–214, 2003. 20

[CMDK+07] S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis. On the Merging of Dung's Argumentation Systems. *Artificial Intelligence*, 171:740–753, 2007. 56

[Con10] V. Conitzer. Making decisions based on the preferences of multiple agents. *Commun. ACM*, 53(3):84–94, 2010. 8

[CP10] M. Caminada and G. Pigozzi. On judgment aggregation in abstract argumentation. *Journal of Autonomous Agents and Multi-Agent Systems*, 2010. 48

[CS02] V. Conitzer and T. Sandholm. Vote elicitation: complexity and strategy-proofness. In *Proceedings of AAAI-02*, pages 392–397, 2002. 20

[CS05] V. Conitzer and T. Sandholm. Communication complexity of common voting rules. In *Proceedings of EC'05*, pages 78–87, 2005. 18, 22

[CSS06] P. Cramton, Y. Shoham, and R. Steinberg, editors. *Combinatorial Auctions*. MIT Press, 2006. 29

[dC85] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris, 1785. 17

[DC06] P. Dunne and Y. Chevaleyre. The complexity of deciding reachability properties of distributed negotiation schemes. *Theoretical Computer Science*, 396(1–3):113–144, 2006. 38

[dKBKZ09] B. de Keijzer, S. Bouveret, T. Klos, and Y. Zhang. On the complexity of efficiency and envy-freeness in fair division of indivisible goods with additive preferences. In F. Rossi and A. Tsoukiàs, editors, *ADT*, volume 5783 of *Lecture Notes in Computer Science*, pages 98–110. Springer, 2009. 34

[DKNS01] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th WWW*, pages 613–622, 2001. 18

[DM02] A. Darwiche and P. Marquis. A knowledge compilation map. *JAIR*, 17:229–264, 2002. 21

[DMA08] Y. Dimopoulos, P. Moraitis, and L. Amgoud. Theoretical and computational properties of preference-based argumentation. In M. Ghallab, C. D. Spyropoulos, N. Fakotakis, and N. M. Avouris, editors, *ECAI*, volume 178 of *Frontiers in Artificial Intelligence and Applications*, pages 463–467. IOS Press, 2008. 46

[Dun95] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-persons games. *Artificial Intelligence*, 77:321–357, 1995. 47

[Dun05] P. E. Dunne. Extremal behaviour in multiagent contract negotiation. *Journal of Artificial Intelligence Research*, 23:41–78, 2005. 43

[DWL05] P. E. Dunne, M. Wooldridge, and M. Laurence. The complexity of contract negotiation. *Artificial Intelligence*, 164(1–2):23–46, 2005. 38

[ECEM06] S. Estivie, Y. Chevaleyre, U. Endriss, and N. Maudet. How equitable is rational negotiation? In *Proc. 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006)*, pages 866–873. ACM Press, 2006. 35, 39

[EFS10] E. Elkind, P. Faliszewski, and A. Slinko. Cloning in elections. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 2010. 25

[EKM07] P. Everaere, S. Konieczny, and P. Marquis. The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*, 28:49–105, 2007. 12

[EM04] U. Endriss and N. Maudet. Welfare engineering for multiagent systems. In A. Omicini, P. Petta, and J. Pitt, editors, *Proceedings of the 4th International Workshop Engineering Societies in the Agent World (ESAW-2003)*, volume 3071 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 93–106. Springer-Verlag, 2004. 39

[EM05] U. Endriss and N. Maudet. On the communication complexity of multilateral trading: Extended report. *Journal of Autonomous Agents and Multiagent Systems*, 11(1):91–107, 2005. 42, 43

[EMST03a] U. Endriss, N. Maudet, F. Sadri, and F. Toni. On optimal outcomes of negotiations over resources. In J. S. Rosenschein, T. Sandholm, M. Wooldridge, and M. Yokoo, editors, *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003)*, pages 177–184. ACM Press, July 2003. 35

[EMST03b] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Protocol conformance for logic-based agents. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003)*, pages 679–684. Morgan Kaufmann Publishers, August 2003. 12

[EMST03c] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Resource allocation in egalitarian agent societies. In A. Herzig, B. Chaib-draa, and Ph. Mathieu, editors, *Secondes Journées Francophones sur les Modèles Formels d'Interaction (MFI-2003)*, pages 101–110. Cépaduès-Éditions, May 2003. 39

[EMST04] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Logic-based agent communication protocols. In F. Dignum, editor, *Advances in agent communication languages*, volume 2922 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 91–107. Springer-Verlag, 2004. 12

[EMST06] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Negotiating socially optimal allocations of resources. *Journal of Artificial Intelligence Research*, 25:315–348, 2006. 36, 39, 42

[ER91] E. Ephrati and J. S. Rosenschein. The Clarke tax as a consensus mechanism among automated agents. In *Proceedings of the 9th AAAI Conference on Artificial Intelligence (AAAI-1991)*, 1991. 18

[FHH09] P. Faliszewski, E. Hemaspaandra, and L. Hemaspaandra. Multimode control attacks on elections. In *Proceedings of IJCAI-09*, pages 128–133, 2009. 25

[Fis73] P. Fishburn. *The Theory of Social Choice*. Princeton University Press, 1973. 19

[FK74] A. Feldman and A. Kirman. Fairness and envy. *The American Economic Review*, 64(6):995–1005, 1974. 34

[Fla96] P. Flach. Abduction and induction: Syllogistic and inferential perspectives. In *Proc. of the Workshop on Abductive and Inductive Reasoning*, LNAI 3259, pages 34–52. Springer, 1996. 51

[FP10] P. Faliszewski and A. D. Procaccia. Ai's war on manipulation: Are we winning? *AI Magazine*, 2010. forthcoming. 18, 20

[Gor96] T. F. Gordon. Computational dialectics. In Peter Hoschka, editor, *Computers as Assistants - A New Generation of Support Systems*, pages 186–203. Lawrence Erlbaum Associates, 1996. 48

[GP05] P. Gradwell and J. A. Padget. Markets vs auctions: Approaches to distributed combinatorial resource scheduling. *Multiagent and Grid Systems*, 1(4):251–262, 2005. 29

[Gra97] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997. 31

[HDM10] N. Hadidi, Y. Dimopoulos, and P. Moraitis. Argumentative alternating offers. In W. van der Hoek, G. A. Kaminka, Y. Lespérance, M. Luck, and S. Sen, editors, *AAMAS*, pages 441–448. IFAAMAS, 2010. 46

[Hud89] O. Hudry. Median linear orders : heuristics and a branch and bound algorithm. *European Journal of Operational Research*, 42(3):313–325, 1989. 17

[Kaw10] R. Kawasaki. Farsighted stability of the competitive allocations in an exchange economy with indivisible goods. *Mathematical Social Sciences*, 59(1):46 – 52, 2010. 33

[KKIA+10] A. Kakas, G. Kern-Isberner, L. Amgoud, N. Maudet, and P. Moraitis. ABA: Argumentation-based agents. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)*, August 2010. Short paper. 12

[KKT93] A. C. Kakas, R. A. Kowalski, and F. Toni. Abductive logic programming. *Journal of Logic and Computation*, 2(6):719–770, 1993. 46

[KKW10] B. Klaus, F. Klijn, and M. Walzl. Farsighted house allocation. *Journal of Mathematical Economics*, In Press, Corrected Proof:–, 2010. 33

[KL05] K. Konczak and J. Lang. Voting procedures with incomplete preferences. In *Proc. IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*, 2005. 20

[KM03] A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of AAMAS03*, 2003. 12

[KMD94] A. Kakas, P. Mancarella, and P. M. Dung. The acceptability semantics for logic programs. In *Proceedings of the International Conference on Logic Programming*, 1994. 12

[KMM05] A. Kakas, N. Maudet, and P. Moraitis. Modular representation of agent interaction rules through argumentation. *Journal of Autonomous Agents and Multiagent Systems*, 11(2):189–206, 2005. Special Issue on Argumentation in Multi-Agent Systems. 12

[KN97] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997. 21, 27

[KP99] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In C. Meinel and S. Tison, editors, *STACS*, volume 1563 of *Lecture Notes in Computer Science*, pages 404–413. Springer, 1999. 14

[KP05] S. Konieczny and R. Pino Prez. Propositional belief base merging or how to merge beliefs/goals coming from several sources and some links with social choice theory. *European Journal of Operational Research*, 160(3):785–802, 2005. 13

[Kra01] S. Kraus. *Strategic Negotiation in Multiagent Environments*. MIT Press, 2001. 7

[KTL+06] S. Koenig, C. A. Tovey, M. G. Lagoudakis, E. Markakis, D. Kempe, P. Keskinocak, A. J. Kleywegt, A. Meyerson, and S. Jain. The power of sequential single-item auctions for agent coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1625–1629. AAAI Press, 2006. 29

[Lan04] J. Lang. Logical preference representation and combinatorial vote. *Annals of Mathematics and Artificial Intelligence*, 42(1):37–71, 2004. 17

[LD03] C. List and J. Dryzek. Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science*, 33(1):1–28, 2003. 45

[LL00] C. Lafage and J. Lang. Logical representation of preferences for group decision making. In *Proceedings of KR2000*, pages 457–468, 2000. 31

[LMS06] D. Lehmann, R. Müller, and T. Sandholm. The winner determination problem. In P. Cramton et al., editors, *Combinatorial Auctions*. MIT Press, 2006. 34

[Lou02] R. Loui. Process and policy: Resource-bounded nondemonstrative reasoning. *Computational Intelligence*, 14(1):1–38, 2002. 48

[Mar00] P. Marquis. Consequence finding algorithms. In Gabbay D. et Smets Ph. (srie eds.) Moral S. et Kohlas J. (eds.), editor, *Handbook on Defeasible Reasoning and Uncertainty Management Systems*, volume 5, chapter 2, pages 41–145. Kluwer Academic Publisher, 2000. 47

[MCd02] N. Maudet and B. Chaib-draa. Commitment-based and dialogue-game based protocols: new trends in agent communication languages. *The Knowledge Engineering Review*, 17(2):157–179, June 2002. 12

[Mou88] H. Moulin. *Axioms of Cooperative Decision Making*. Cambridge University Press, 1988. 33

[MPRJ10] R. Meir, M. Polukarov, J. S. Rosenschein, and N. R. Jennings. Convergence to equilibria in plurality voting. In M. Fox and D. Poole, editors, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2010. 18

[MRBC08] J. Ma, A. Russo, K. Broda, and K. Clark. Dare: a system for distributed abductive reasoning. *Journal of Autonomous Agents and Multiagent Systems (JAAMAS)*, 16(3):271–297, 2008. 48

[NIIR10] H. Nabeshima, K. Iwanuma, K. Inoue, and O. Ray. SOLAR: An automated deduction system for consequence finding. *AI Communications*, 23:183–203, 2010. 47

[Nis06] N. Nisan. Bidding languages for combinatorial auctions. In P. Cramton et al., editors, *Combinatorial Auctions*. MIT Press, 2006. 31

[OMT07] W. Ouerdane, N. Maudet, and A. Tsoukias. Arguing over actions that involve multiple criteria: A critical review. In K. Mellouli, editor, *Proceedings of the Ninth European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty (ECSQARU–2007)*, volume 4724 of *LNCS*, pages 308–319. Springer-Verlag, October 2007. 27

[OMT08] W. Ouerdane, N. Maudet, and A. Tsoukias. Argument schemes and critical questions for decision aiding process. In Philippe Besnard, Sylvie Doutre, and Anthony Hunter, editors, *Computational Models of Argument: Proceedings of COMMA 2008*, Frontiers in Artificial Intelligence and Applications, pages 285–296. IOS Press, 2008. 27

[PHG00] D. M. Pennock, E. Horvitz, and C. L. Giles. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *AAAI/IAAI*, pages 729–734. AAAI Press / The MIT Press, 2000. 18

[Poo89] D. Poole. Explanation and prediction: An architecture for default and abductive reasoning. *Computational Intelligence*, 5(2):97–110, 1989. 46, 53

[PR07] A. Procaccia and J. Rosenschein. Average-case tractability of manipulation in elections via the fraction of manipulators. In *Proceedings of the 6th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-07)*, pages 718–720, 2007. 23

[Pra05] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 15:1009–1040, 2005. 45, 48, 56

[PRVW07] M. S. Pini, F. Rossi, K. Venable, and T. Walsh. Incompleteness and incomparability in preference aggregation. In *Proceedings of IJCAI-2007*, pages 1464–1469, 2007. 20

[PS97] H. Prakken and G. Sartor. Argument-based logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75, 1997. 48

[PSJ98] S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8:261–192, 1998. 48

[PWA03] S. Parsons, M. Wooldridge, and L. Amgoud. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation*, 13(3):347–376, 2003. 48

[Raw71] J. Rawls. *A Theory of Justice*. Oxford University Press, 1971. 34

[RL09] I. Rahwan and K. Larson. Argumentation and game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009. 48

[RPH98] M. H. Rothkopf, A. Pekeč, and R. M. Harstad. Computationally manageable combinational auctions. *Management Science*, 44(8):1131–1147, 1998. 34

[RRJ+03] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(4):343–375, 2003. 46

[RS09] I. Rahwan and G. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009. 47

[RT10] I. Rahwan and F. Tohmé. Collective argument evaluation as judgement aggregation. In *Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'09)*, 2010. 56

[RZ94] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter*. MIT Press, 1994. 7, 8, 11

[San98] T. W. Sandholm. Contract types for satisficing task allocation: I Theoretical results. In *Proc. AAAI Spring Symposium: Satisficing Models*, 1998. 33, 35

[SLB09] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoetic, and Logical Foundations*. Cambridge University Press, 2009. 7, 29

[Smi80] R. Smith. The contract net protocol: high level communication and control in distributed problem solver. *IEEE Transcations on Computers*, 29:1104–1113, 1980. 29

[SS07] S. Saha and S. Sen. An efficient protocol for negotiation over multiple indivisible resources. In *Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI-2007)*, pages 1494–1499, 2007. 30

[STT01] F. Sadri, F. Toni, and P. Torroni. Dialogues for negotiation: agent varieties and dialogue sequences. In J. J. Meyer and M. Tambe, editors, *Intelligent Agent Series VIII: Proceedings of the 8th International Workshop on Agent Theories, Architectures, and Languages (ATAL 2001)*, volume 2333 of *Lecture Notes in Computer Science*, pages 69–84. Springer Verlag, Berlin, Germany, 2001. 12

[UCEL09] J. Uckelman, Y. Chevaleyre, U. Endriss, and J. Lang. Representing utility functions via weighted goals. *Mathematical Logic Quarterly*, 55(4):341–361, 2009. 31

[Wal08] T. Walsh. Complexity of terminating preference elicitation. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS-08)*, pages 967–974, 2008. 20

[Wel96] M. P. Wellman. Market-oriented programming: some early lessons. In S. Clearwater, editor, *Market-based Control: A paradigm for Distributed Resource Allocation*. World Scientific Publishing, 1996. 29

[Woo00] M. Wooldridge. Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1):9–31, 2000. 12

[Woo09] M. Wooldridge. *An Introduction to Multiagent Systems*. John Wiley and Sons, 2009. Second Edition. 7

[XC08a] L. Xia and V. Conitzer. Determining possible and necessary winners under common voting rules given partial orders. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-08)*, pages 196–201, 2008. To appear. 20

[XC08b] L. Xia and V. Conitzer. A sufficient condition for voting rules to be frequently manipulable. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC-08)*, pages 99–108, 2008. 20, 23

[XC09] L. Xia and V. Conitzer. Finite local consistency characterizes generalized scoring rules. In C. Boutilier, editor, *IJCAI*, pages 336–341, 2009. 20

[XC10a] L. Xia and V. Conitzer. Compilation complexity of common voting rules. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 2010. To appear. 22

[XC10b] L. Xia and V. Conitzer. Stackelberg voting games: Computational aspects and paradoxes. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-2010)*, 2010. 22, 23

[XLM10] L. Xia, J. Lang, and J. Monnot. Possible winners when new candidates are added: new results coming up! In *Proceedings of the Third Workshop on Computational Social Choice (COMSOC-2010)*, 2010. 26

[ZPR09] M. Zuckerman, A. D. Procaccia, and J. S. Rosenschein. Algorithms for the coalitional manipulation problem. *Artificial Intelligence*, 173(2):392–412, 2009. 23