Artificial Intelligence Review **21:** 3–24, 2004.

Rights and Argumentation in Open Multi-Agent Systems

EDUARDO ALONSO
Department of Computing, City University, London EC1V 0HB, UK
(Fax: +44 20 7040 8587; E-mail: eduardo@soi.city.ac.uk)

Abstract. As utility calculus cannot account for an important part of agents' behaviour in Multi-Agent Systems, researchers have progressively adopted a more normative approach. Unfortunately, social laws have turned out to be too restrictive in real-life domains where autonomous agents' activity cannot be completely specified in advance. It seems that a halfway concept between anarchic and off-line constrained interaction is needed. We think that the concept of right suits this idea. Rights improve coordination and facilitate social action in Multi-Agent domains. Rights allow the agents enough freedom, and at the same time constrain them (prohibiting specific actions). Besides, rights can be understood as the basic concept underneath open normative systems where the agents reason about the code they must abide by. Typically, in such systems this code is underspecified. On the other hand, the agents might not have complete knowledge about the rules governing their interaction. Conflict situations arise, thus, when the agents have different points of view as to how to apply the code. We have extended Parsons's et al. argumentation protocol (Parsons et al. 1998a, b) to normative systems to deal with this problem.

Keywords: argumentation, coordination, multi-agent systems, rights, social action

1. Introduction

So far, the Rational Choice Theory (RCT) has been the most influential theory for designing agents in Artificial Intelligence and Distributed Artificial Intelligence. According to this approach to rationality, agents with complete knowledge make their decisions in order to maximise their own utilities. In this traditional, non-constrained approach agents have been assumed "free": They act of their own accord and are not subject to any set of (social) rules. However fruitful this approach has been, there have been pointed out (e.g., Reiner 1995) several drawbacks in RCT, namely:

1. In real dynamic domains agents do not have enough information or time to perform complex, optimal utility calculus. An agent does not know all the alternatives, does not know the exact outcome of each, and does not have a complete preference order for those outcomes. This problem becomes particularly grave in Multi-Agent Systems (MAS) due to the presence of various agents, each with their own beliefs, goals and intentions.

Table 1. RCT and social action. The standard form of presentation of the rules of two-player simultaneous games in RCT is a matrix, with elements representing pay-offs. In our example, *Agent 1* has three possible courses of action, *A*, *B*, and *C*; *Agent 2*, on the other hand, has two choices, *L* and *R*. The matrix tells us the pay-offs of each combination of actions for both agents. For instance, the result of *Agent 1* executing *B* and *Agent 2* executing *R*, that is, the result of *(R,B)*, will be assessed as 2 in *Agent 1*'s scale of preferences and 1 in *Agent 2*'s.

		Agent 1		
		A	B	C
Agent 2	L	3,4	**2,5**	1,3
	R	**4,8**	1,2	0,9

2. On the other hand, the utilitarian approach has failed in explaining cooperation and social action. As illustrated in the *Prisoner's Dilemma*, agents can choose dominant, but socially irrational strategies. In the example depicted in Table 1, the matrix tells us that *(R,A)* is a rational strategy to be followed by both agents. What is more, is the Pareto solution to the game, that is, this solution (4,8) cannot be improved upon for one agent without lowering the other agent's utility. We would, thus, expect agents to cooperate and agree on this joint strategy. This is not the case, though, as *(R,A)* is not a stable strategy: At least one of the agents has an incentive to deviate from it. Assuming that *Agent 2* chooses *R*, then *Agent 1* will be tempted to choose *C*, as *(C,R)* gives it 9. *Agent 2* will have known that when making its choice and, consequently, will have chosen *L*, as it prefers *(L,C)*, that is, (1,3), over *(R,C)*, that is, over (0,9). Now, *Agent 1* will also have predicted what *Agent 2* was thinking and gone for *B* as *(L,B)* is preferred over *(L,C)*. At this point, none of the agents will have any reason for changing their strategies. *(L,B)* is in equilibrium. So the equilibrium point is (2,5), even though (4,8), the Pareto solution, is more attractive. Agents face a "trust dilemma": They can take a position that, if rational, at least one of them may be tempted not to implement.

   Following this line of argumentation, we cannot explain either collective action or cooperation. There is no notion of social action as a jointly planned course of action: Agents calculate individually and separately their best options. Moreover, communication or negotiation would not help, for agents cannot trust each other and will back down on the agreed commitments. In a word, cooperation is futile.

In order to cope with these problems, the MAS community has adopted a more constrained approach to rationality including conventions, norms and/or social laws. It is well-known that agents working under norms do not need to calculate continuously their utilities and, consequently, do not need complete information. Agents are supposed to act in a somehow predetermined way according to the principle of "mutual expectation". Besides, norms imply that the agents respect certain social constraints that deter them from breaking agreements. Unfortunately, research in this field has fallen into two extreme positions: Shoham and Tennenhlotz (Shoham and Tennenhlotz 1992) have studied off-line social laws, which agents must comply with automatically. Agents are assumed to follow rules just because they are designed to do so. Agents are not seen as autonomous any more. Proposals so formulated are thus closer to Distributed Problem Solving than to MAS.

Alternatively, conventions (e.g., Walker and Wooldridge 1995) have been introduced as rules emerging during repeated encounters in open normative systems. The problem here is that no notion of sanction is considered. Consequently, if the agents have the chance to calculate their utility each time they interact, conventions are continually under consideration. In other words, following a convention is not always a stable strategy.

It seems, therefore, that we need a concept that allows agents to reason and make decisions, but that implies enforcement at the same time. That is, we need a halfway concept (neither off-line nor strictly on-line) that guides, but does not control, the behaviour of autonomous agents. Our contention is that the concept of "right" suits these requirements.

One of the consequences of providing autonomous agents with rights is that coordination is not a "follow-the-rule" process. In real-life normative situations, the code to which the agents are referred to is typically incomplete and/or ambiguous. To avoid conflict in such cases, agents use to argue about who has the right to do what.

The remainder of the paper is structured as follows. In the second section, we present the concept of rights as liberties; in the third section, a simple theory of rights is characterised; in the fourth section, we consider what we can gain by introducing rights in the coordination process in terms of complexity, efficiency, stability, and flexibility; in the fifth section, the relationships between constrained and unrestricted behaviour in the coordination process are studied; in the sixth section, argumentation is introduced in normative systems to solve problems due to underspecification or ambiguity; we shall finish with some conclusions and further research.

2. Rights

Roughly stated, a right is considered as a set of restrictions on the agents' activities which allow them enough freedom, but at the same time constrain them. Not surprisingly, some authors (e.g., Tennenhlotz 1998) have expressed the same idea from a RCT perspective, by introducing some constraints in the set of strategies available to the agents. In so doing, agents are free to converge on "stable social laws" (qualitative equilibria). However interesting this approach may be, it presents a serious handicap: To make sure that the agents choose a stable and efficient strategy, the designer decides beforehand which strategies should be eliminated. The designer, therefore, manipulates the process and creates an "illusion of freedom".

We interpret this concept from a social approach, as was advanced in (Alonso 2002): To explain social behaviour we need to think of the agent as a *homo sociologicus* rather than as a *homo economicus*.

Generally speaking, if an agent has the right to execute a set of actions then (a) it is permitted to perform it (under certain constraints or obligations), (b) the rest of the group is not allowed to execute any action inhibiting the agent from exercising its right, and (c) the group is obliged to prevent this inhibitory action.

We can illustrate this idea with a simple traffic example that will be used throughout the paper. In Figure 1, x has the right to drive along the main A road under certain constraints (have the corresponding licence, respect the speed limit, drive on the left, etc.); y is not allowed to take this road from the B road at that junction at the same time, because this action inhibits x's right;[1] finally, the rest of the group must stop y from breaking the law and, if needed, punish the offence. In large organizations, the group can delegate these responsibilities to expert agents, police-traffic agents in this case.

This third point follows from Castelfranchi's *right to claim* (Castelfranchi 1995), according to which any agent has the right to ask for help if his counterpart in the interaction (a short term deal or a long term socially established pattern of behaviour) does not abide by the terms of the contract. We extend, nonetheless, this notion and talk about the *right to be protected*: Agents have the right to be aided even when they themselves do not know that their rights are under threat (and therefore do not make any claim). So, even if x does not know that y has the intention of breaking the law and does not claim the group for help (right to claim), the group is obliged to assist him (right to be protected).

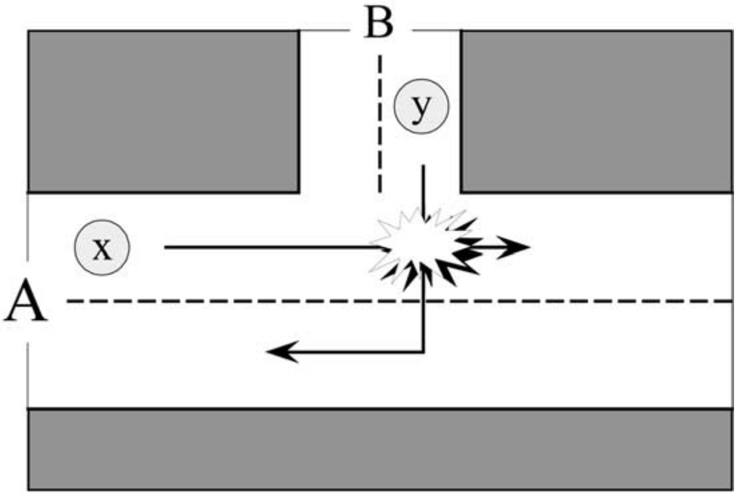

Figure 1. A traffic example. Agent *x* is driving along road *A* and agent *y* is trying to drive from *B* to *A* at the same time. Without a theory of rights, *x* and *y* will crash against each other as they have conflicting goals and no way of solving the problem.

2.1. *Three facts about rights*

We are introducing now three distinctive facts about rights. These rights are related very closely to, but are not reducible to permissions or the right to claim. They are very basic rights, represent universal interests and form systems.

1. **Liberties:** Unlike in (Norman et al. 1998), rights are not considered in this paper as permissions. We are interested in basic rights, in what (Barry 1989) called liberties. Of course, once a right is adopted, it works like a permission. The main difference between rights and permissions lies in the fact that rights are universal statements: (a) the entire group agrees on them; (b) they establish equality; (c) they apply for the long-term. Nobody can delegate or trade with rights. Permissions, on the other hand, are relative.

   Let's illustrate these differences with an example. In many countries, drivers must pay a fixed amount of money to have the permission to drive along a motorway. Once a driver has paid, he has the right to drive along the motorway. We can think of such a condition as the need for a permission: There is an agreement saying that if an agent pays, say, £5 then he gets the permission to use the road. However, this agreement is quite special, because it is applicable to any agent anytime. Agents do not have to ask for permission each time they want to drive along the motorway.

That is not the case with mere premissions. A driver can reach an agreement with the owner of a private parking: To pay £30 per night. That is a bilateral one-shot agreement. If another driver wants to use the parking, it has no rights until a new agreement with the parking owner is reached. The previous agreement does not set up a precedent. Renegotiation is required.

Rights are essentially social concepts: the notion of group is a guarantee that agents' rights (and their corresponding obligations) are observed through sanctions and compensations. Moreover, rights mean that all the agents in a group are closely related: Those agents exercising a right (active agents) are constrained by the obligations linked to that right, and the others (passive agents) are constrained by prohibitions (it is forbidden to violate others' rights).

2. **Hierarchies:** Rights form systems. Permissions do not. As a consequence, rights cannot be adopted individually, one by one. For example, I am not allowed to accept the right to drive on the left (because it is instrumental in satisfying my goal of driving work on time) and reject at the same time the right to live (which may turn out to be in conflict with my goal, so I could be tempted to run over any pedestrian crossing the road in my way).

   Ideally, rights are totally ordered in a hierarchy according to the values they represent. The higher a right is in the hierarchy, the more important the value it represents. Therefore, a right is only overruled by another in a higher position. In theory, conflicts between rights are automatically solved according to this order, because higher rights act as constraints for lower rights. As driving along *A* has precedence over driving from *B* to *A*, one of the constraints for exercising this second right is that there is no agent exercising the first.

3. **Interests vs Utilities:** The main purpose of recognising rights is to protect certain interests of individuals against additive calculations of utilities. Utilitarians can reply that the disutility of frustrating expectations and the utilities of honouring them must be introduced in the utility equations.[2] On the contrary, rights are beyond utility calculus: They represent values or interests. The point of speaking of rights is precisely that we do not want such utilities calculated. Quoting Nielsen and Shiner (Nielsen and Shiner 1977)

   > "For the sake of security and psychological stability we want interests of individuals (...) protected against such calculations, and hence we grant them rights (...). And where rights exist, they are not to be overridden by mere utilities" (Nielsen and Shiner 1977, p. 127).

An agent is entitled to exercise a right (in a way legitimised in the stipulation of the right) despite the greater utility realised by breaking this right. In the traffic example, it doesn't matter if there are three agents y, w, and z intending to drive from B to A, so that the addition of the utilities that these three agents would get by breaking the law would be greater than the utility that x gets by driving along A. This kind of calculations are out of the question, precisely because of x's right to drive along A.

A direct implication of defining rights over utilities is that agents can exercise their rights even though their behaviour is considered wrong. It is important to differentiate between "legally" right and "morally" right. It can be "morally" wrong to exercise a "legal" right: Nobody is allowed to prevent Sunday drivers from driving as long as they abide by the traffic code. They just drive badly.

3. A Language for Describing Rights

This section presents a formal characterisation of the concept of right. Apart from different axioms for characterising rights and permissions respectively, our model follows Norman's et al. (Norman et al. 1998). Readers are referred to that paper for a complete description of the syntax and semantics of the proposed language.

3.1. *Syntax*

The language $\mathcal{L}$ is based on dynamic logic, because we want to talk about agents performing actions, action sequences, etc. We will have three basic sets: propositional variables, $\mathtt{P}$, agents, $\mathtt{Agents}$, and actions, $\mathtt{Actions}$. $\mathtt{Actions}$ is completely ordered according to their social relevance, $(\mathtt{Actions}, \leq)$. The symbols for our language of rights, $\mathcal{L}$, are as follows: $\mathtt{x}$ and $\mathtt{y}$ denote agents. The *group*, an agent after all, will be represented by the symbol $\mathtt{g}$. α, β, γ, and δ denote individual actions.

The main predicates in our language refer to agents' mental attitudes, action expressions and normative expressions. Unlike in (Krogh 1996; Staniford 1994) we are not using deontic logic to express deontic notions. This is because non-forbidden actions are not necessarily allowed actions. In our model, rights must be explicitly established.

The fact that an agent $\mathtt{x}$ has the intention to execute action α, is represented as $\mathtt{I}(\mathtt{x}, \alpha)$. $\mathtt{Done}(\mathtt{x}, \alpha)$ is used to denote that agent $\mathtt{x}$ has just performed action α. $\mathtt{Happens}(\alpha)$ means that α does occur.

As for normative notions, that $\mathtt{x}$ is allowed to execute α is represented as $\mathtt{A}(\mathtt{x}, \alpha)$; if it is forbidden to execute α, we will use $\mathtt{F}(\mathtt{x}, \alpha)$; finally, if there is an obligation, $\mathtt{O}(\mathtt{x}, \alpha)$ will be used. The most important predicate refers to the concept of right: $\mathtt{R}(\mathtt{x}, \alpha)$ means that agent x has the right to execute α. The set of potential rights, Rights, are ordered according to the actions they stipulate.

Atomic propositions and compound formulae of $\mathcal{L}$ are defined as usual. $\mathtt{Inh}(\alpha, \beta)$ means that α inhibits β: If α happens then β does not happen. Formally, $\mathtt{Inh}(\alpha, \beta)$ iff $\mathtt{Happens}(\alpha) \rightarrow \neg\mathtt{Happens}(\alpha; \beta?)$, where $\alpha; \beta$ means “do α followed by β”, and ϕ? means “proceed if ϕ is true”.

3.2. *Semantics*

The semantics for the language of rights, $\mathcal{L}$, is based on a possible worlds model (Hintikka 1962), *à la* Norman et al. (Norman et al. 1998). The class of models of $\mathcal{L}$ that we are interested in are those satisfying the constraints introduced by the following axioms.

3.3. *Axiomatics*

It remains to provide the axiomatics for $\mathcal{L}$.

- **Permission:** Firstly, to have a right does not automatically allow the right-holder to exercise its content. There are some conditions with which the agents have to comply. That is, that an agent has the right to execute an action does not mean that it is legal for him to execute such an action.

$$\neg\exists \mathtt{y}(\mathtt{R}(\mathtt{x}, \alpha) \wedge \mathtt{R}(\mathtt{y}, \beta) \wedge \mathtt{I}(\mathtt{y}, \beta) \wedge \alpha < \beta \wedge \mathtt{Inh}(\beta, \alpha)) \rightarrow \mathtt{A}(\mathtt{x}, \alpha) \quad (1)$$

It is one thing to be allowed to exercise a right; it is another to have the intention of exercising it. Rights do not elicit actions. Social commitments do (Jennings 1993). If the agent is allowed to execute an action, he has the chance to choose whether or not to proceed. So, rights provide the agents with freedom, for they depend on their own motivation to make a decision and act. On the other hand, having an intention does not entitle the agent to execute the corresponding action in normative systems: In normative scenarios, the agent must have the legal capability (it has to be allowed) to do so.

- **Prohibition:** If an agent is allowed to execute an action and has the intention to do so, then no other agent is allowed to exercise a lower inhibiting right. In the traffic example, *y* is allowed to exercise the right of taking the junction *A–B*, as long as *x* does not have the intention to drive along the main road.

$$(\mathtt{A}(\mathtt{x}, \alpha) \wedge \mathtt{I}(\mathtt{x}, \alpha) \wedge \mathtt{Inh}(\beta, \alpha)) \rightarrow \mathtt{F}(\mathtt{y}, \beta) \quad (2)$$

For simplicity, we have adopted in this paper a relativistic approach in which for an action to be forbidden it has to inhibit someone's rights. We could, however, have introduced a more general notion of illegality: It is forbidden to drive at more than 50 mph in town, regardless of whether or not an agent driving at, say, 60 mph is inhibiting someone else's rights. In such a case, nobody has to be protected. Nevertheless, the group has still the obligation to stop the offender.

- **Obligation:** To prevent and/or to sanction.

  Prevention: If an agent has the intention of executing a banned action, β, then the group is obliged to accomplish an inhibitory action and prevent the crime before β is done.

  $$(\mathtt{F}(\mathtt{y}, \beta) \wedge \mathtt{I}(\mathtt{y}, \beta) \wedge \mathtt{Inh}(\gamma, \beta)) \rightarrow \mathtt{O}(\mathtt{g}, \gamma) \tag{3}$$

  Sanction: Finally, if that action has been executed, then the group has to sanction the offender by inhibiting some of its rights (e.g. suspending his licence).

  $$(\mathtt{F}(\mathtt{y}, \beta) \wedge \mathtt{Done}(\mathtt{y}, \beta) \wedge \mathtt{R}(\mathtt{y}, \delta) \wedge \mathtt{Inh}(\gamma, \delta) \rightarrow \mathtt{O}(\mathtt{g}, \gamma) \tag{4}$$

  Obviously, (3) and (4) refer to x's right to be protected. If the group has to prevent the offence, it has to be endowed with an efficient mechanism to recognise intentions. On the other hand, if the crime is eventually committed, the group has to know if it was intentional. Different sanctions correspond to different degrees of intentionality (for instance, murder is more severely punished than manslaughter). It is true that justice is concerned about the legality of actions. But it is also true that when it comes to do justice, agents' intentions and beliefs have to be taken into account. If I drive over a pedestrian, no doubt I have executed an illegal action. However, if the pedestrian was jaywalking and run unexpectedly into my car, then it would be unfair to blame me for the accident. We will issue intention recognition in future papers.

More specifically, we can explain what to exercise a right means following the sample displayed in Table 2.

We have omitted a few features in this figure: Firstly, x has to observe some constraints if it decides to execute α. If it does not do so, then y has the right to claim; secondly, when an agent has performed a forbidden action, the offended agent is usually more concerned about compensations than about sanctioning the offender. Therefore, actions restoring (part of) their rights have to be added to the algorithm; finally, sanctions should be introduced in preventive cases (like in attempted murder), not only if the prohibited action is eventually carried out.

Table 2. Axioms of the system of rights. We assume that (a) x is allowed to execute α (**R1**), (b) y has the right to execute β and δ, (c) β inhibits α, and γ inhibits β and δ, (d) α has priority over β. Under such circumstances, **R2** states that if x has the intention of executing α then y is forbidden to execute β; **R3** reads that if, ignoring **R2**, y tries to execute β, then the group should prevent it by executing an inhibitory action; finally, **R4** contemplates the possibility that y had already performed β, in which case a sanction inhibiting y's rights must be implemented.

CONDITIONS					
	• $\mathtt{R}(\mathtt{x},\alpha)$, $\mathtt{R}(\mathtt{y},\beta)$ and $\mathtt{R}(\mathtt{y},\delta)$ • $\mathtt{Inh}(\beta,\alpha)$, $\mathtt{Inh}(\gamma,\beta)$, and $\mathtt{Inh}(\gamma,\delta)$ • $\alpha > \beta$				
RULES					
R1	$\mathtt{A}(\mathtt{x},\alpha)$				
R2	**IF**	$\mathtt{I}(\mathtt{x},\alpha)$	**THEN**	$\mathtt{F}(\mathtt{y},\beta)$	
R3	**IF**	$\mathtt{I}(\mathtt{y},\beta)$	**THEN**	$\mathtt{O}(\mathtt{g},\gamma)$	*Prevention*
R4	**IF**	$\mathtt{Done}(\mathtt{y},\beta)$	**THEN**	$\mathtt{O}(\mathtt{g},\gamma)$	*Sanction*

4. What Do We Gain by Using Rights?

We contend that the idea of using rights is worthy of consideration because it makes easier to have agents coordinated. In seeking for argue this hypothesis we present a qualitative (rather than quantitative) analysis. As it has been repeatedly pointed out (e.g., O'Hare and Jennings 1996; Sycara 1998; Weiss 1999), coordination is mainly concerned with complexity, efficiency, stability, and flexibility. Roughly speaking, complexity refers to how difficult it is to find a solution, and depends on the amount of information and/or time required to represent and solve the problem; efficiency speaks about the quality of the outcome, how good it is; then, the solution must be stable, that is, agents should have no reason to diverge from it; and finally, for a MA system to be flexible means that the agents are able to respond by themselves to the changing environment. In dynamic MA systems, we want autonomous agents to obtain the best stable results using as few resources as possible. We illustrate in the following Table 3. how different approaches work to get agents coordinated, and what we gain by using rights:

1. **Complexity:** We can see rights as social conditions to execute actions. An agent must have the legal capability to perform an action. Consequently,

Table 3. Coordination in different MAS approaches. RCT, closed normative systems, and rights (or open normative systems) are intuitively evaluated according to their complexity, efficiency, stability and flexibility.

	RCT	Norms	Rights
Complexity	High	Medium	Low
Efficiency	Low	High	High
Stability	High	High	High
Flexibility	High	Low	High

- the representational complexity is reduced. Theoretically, x does not need to know from, to where, when or how other agents are driving in Figure 1. Agents do not have to anticipate all possible course of events;
- performance itself is also improved: The conditions for success are partially assured through prohibitions and obligations. Rights restrict potentially harmful interactions, and avoid conflict by cutting some paths;
- moreover, rights reduce negotiation and communication costs. Obviously, x and y do not negotiate each time they meet in the junction A-B.[3]

We depict the traffic problem decision tree in Figure 2, where a means that x is driving along A and b that y is driving from B to A. The first agent has priority, so the first and the third branches are pruned. Agents do not have to reflect on them and reckon what they would individually gain or lose by trying an alternative path. The second branch is executed directly. If compared with other approaches, we maintain low levels of complexity by using rights. RCT is highly complex when it comes to find a solution: Agents have to take into consideration all possible options. As for norms, it may seem that they cope with complexity as rights do: They, too, cut different paths, create safety areas, and reduce communication after all. However, it is worth noticing that it is the designer who establishes off-line the strategy to be followed. So, even though the overall performance is satisfactory, it requires lots of representational work.

2. **Efficiency and stability:** Rights are essentially social notions. The group guarantees through sanctions that agents' rights (and their corresponding obligations) are observed.

   In RCT, agents must sacrifice efficiency to preserve stability. When norms are involved, stability and efficiency (usually in terms of global utility) are assured, but only because agents obey orders. In our approach,

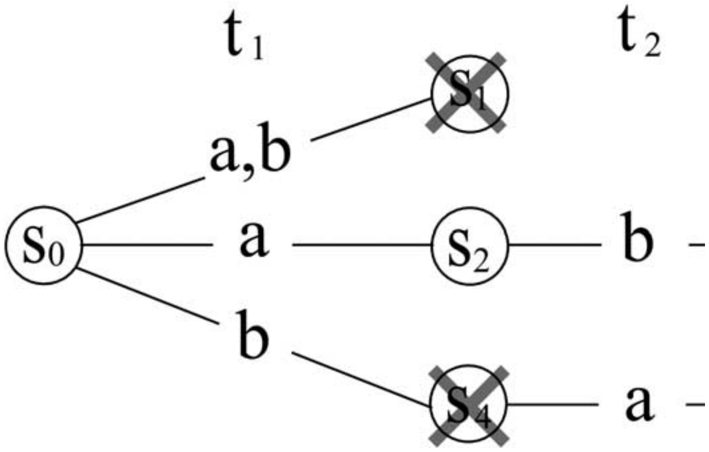

Figure 2. Rights and complexity. *a* represents that agent *x* is driving along *A* and *b* that *y* is driving from *B* to *A* in Figure 1. t_1 and t_2 represent two consecutive time points. S_0 is a description of the initial state of the world. S_i represents the state of the world after the actions linking S_{i-1} to S_i have been executed. For example, S_2 will result as a consequence of performing *a* at S_0 at t_1. By using rights, complexity is reduced. As agent *x* has the right to drive along *A*, neither *a,b* nor *b* can be executed at t_1. Consequently, neither S_1 nor S_4 will be considered.

autonomous agents are free to adopt efficient solutions, for the group is responsible for stability. Stability is now a social concept, neither a strategic nor a normative one. The ‘trust dilemma’ is, therefore, solved: Agents do not have to watch each other; they can cooperate and make joint decisions.

The right to be protected is the main responsible for this dramatic change. With this meta-right at hand, agents would choose (4,8) in Table 1: If *Agent 1* does not abide by this agreement and finally executes *C*, then *Agent 2* is entitled to ask the group to sanction the first agent and force it to restore its rights. In order to assure that the agents will abide by the rules, “Draconian laws” can be introduced.

In the long term, rights introduce fairness. As agents do not know beforehand which role they are going to play in the future, they assess the situation as “Kantian” impartial judges. In the traffic world, agents have to answer this general question: “Is it instrumental to drive on the left to avoid conflict?” No individual parameters are taken into account in answering this question.

3. **Flexibility:** Rights give the agents the chance to decide to execute a set of actions. Right-holders are not committed to any specific action. However, if an agent exercises a right then it is committed to do so under certain constraints. Rights are not procedural, but they create attitudes in the

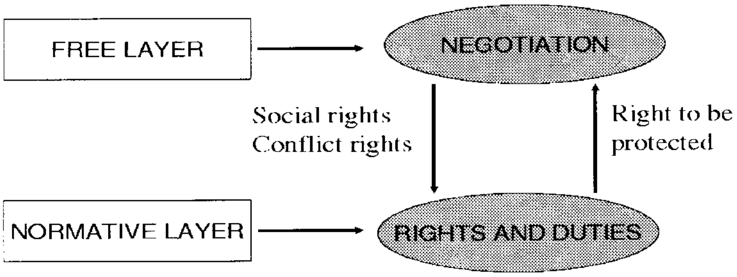

Figure 3. The coordination cycle. Co-ordination is achieved either at a normative layer or at a "free" layer. The first refers to long-term interactions and is governed by rights and duties, while the latter functions on the short term through non-constrained negotiation. Both layers interact though, completing a cycle: Short-term interactions need a guarantee that the deals the agents agree upon will be abided by. The right to be protected, that is, the right to make the group respect the deals, will assure that such layer does work and, thus, that coordination is efficient. On the other hand, social rules do not cover all possible interactions at a normative level. In such cases, agents should be free to reach agreements on how such norms should be implemented.

agents. These agents are not mere vehicles of established norms, but they can decide to abide by or break their obligations. This property is very valuable, because it puts the accent on agents' autonomy, unlike social laws or *ad hoc* binding agreements (Kraus et al. 1995; Rosenschein and Zlotkin 1994).

In so doing, we can establish a clear distinction between rights and social commitments: Social commitments elicit actions, rights do not.

5. Rights in the Coordination Process

To sum up: Rights protect interests, reduce representational as well as procedural complexity, provide the agents with control mechanisms (the right to be protected) to assure stability and efficiency both in short and long term encounters, and preserve autonomy and flexibility.

Yet, a theory of rights alone cannot account for coordination in MAS. Not all interactions are ruled by rights. As is shown in Figure 3, coordination can be achieved through negotiation (or other non-normative coordination mechanism) and/or following rights. The first method rules the agent's "free layer", and the second the "normative layer". The free layer governs unconstrained short term behaviour according to agents' preferences and dependence relationships (Alonso 1998, 1999; Castelfranchi et al. 1992; Sichman et al. 1994; Wooldridge and Jennings 1994), whereas the normative layer guides the long term activity.

All in all, coordination results from the bidirectional interaction between these two layers: Unconstrained behaviour and negotiation (its solver) are supervised by the right to be protected, and negotiation is sometimes necessary to complement the normative layer.

- **From rights to negotiation:** As it has already been mentioned, the right to be protected does not apply only to the execution of rights (when an agent does not abide by the constraints or obligations linked to the right in progress, or when someone tries to inhibit others' rights by executing forbidden actions). It is also applied to one-shot deals: When an agent does not fulfil its commitments, then the right to be protected is exercised by the other part. This right compels the group to force the offender to comply with the agreement.
- **From negotiation to rights:** Conflicts can arise in the normative layer given underspecification. Usually, agents cannot be referred to a complete, unambiguous traffic code. Imagine that two drivers are trying to park in the same place, one backwards, the other forwards. As the right to park does not specify which agent has precedence, conflict follows.[4]

  Experimentation can help to detect and avoid conflict cases (in the parking example, we can introduce a constraint on the right to park saying "it is forbidden to park forwards" or give priority to the right "parking backwards"). However, the dynamic nature of social interaction makes it impossible to write down a perfect system of rights. Moreover, it is very difficult for (informationally) limited agents to know exactly how legal systems work.

Of course, we could adopt a centralized approach and endow a "big brother" with the power to dictate a solution in such circumstances. However, as we are dealing with autonomous agents, we have left in these agents' hands the responsibility for deciding through arguing what to do when it is not clear which norm should be followed. The group watches the process though. That means, first, that potential conflicts can be solved directly by an expert, and, second, that the agents can always ask the group to take part as a judge in the argumentation process.

6. Argumentation

So, agents have to negotiate after all. In order to cope with this need to negotiate in open normative systems such as our traffic world, we are introducing a normative-tailored version of argumentation. Recently, several studies on argumentation in negotiation have been presented as a powerful technique to cooperate and solve conflict situations (Kraus et al. 1998; Parsons et al.

1998a, b; Sierra et al. 1998). In this type of negotiation agents exchange not only proposals and counterproposals, but reasons supporting them. We are going to use Parsons's et al. multi-context argumentative framework. If the reader is familiar with this work, pass to the next subsection. Formally,

Definition 1 *Given an agent* $\mathtt{x}$*, an* argument *for a formula* ϕ *is a pair* $(\phi, \mathtt{P})$ *where* $\mathtt{P} = \{\mathtt{s}_1, \ldots, \mathtt{s}_n\}$ *and either* $\mathtt{s}_\mathtt{i}$ *is a formula in the theories of agent* $\mathtt{x}$ *or* $\mathtt{s}_\mathtt{i} = \Gamma_\mathtt{i} \vdash_{\mathtt{d}_\mathtt{i}} \psi$ *and* $\mathtt{p}_\mathtt{j} \in \Gamma_\mathtt{i}$ *is either a formula in the theories of* $\mathtt{x}$ *or the conclusion of a previous step in* $\mathtt{P}$*, and* ψ *is a formula in the language of* $\mathtt{x}$*, and* $\mathtt{s}_\mathtt{n} = \Gamma_\mathtt{n} \vdash_{\mathtt{d}_\mathtt{n}} \psi$*.*

Definition 2 *An argument* $(\phi, \mathtt{P})$ *is* consistent *if there is no* $\mathtt{s}_\mathtt{i}, \mathtt{s}_\mathtt{j} \in \mathtt{P}$ *such that* $\mathtt{s}_\mathtt{i} = \Gamma_\mathtt{i} \vdash_{\mathtt{d}_\mathtt{i}} \psi$ *and* $\mathtt{s}_\mathtt{j} = \Gamma_\mathtt{j} \vdash_{\mathtt{d}_\mathtt{j}} \neg\psi$*.*

Definition 3 *An argument* $(\phi, \mathtt{P})$ *is* non-trivial *if it is consistent.*

Definition 4 *An argument* $(\phi, \mathtt{P})$ *is* tautological *if all deductive steps in* $\mathtt{P}$ *are built using only rules of inference, bridge rules and axioms of the logics of the agent's units.*

Because in argumentation a proof for a formula only suggests that the formula may be true (rather than indicating that it is true), we can have arguments for and against the same formula. In particular, given an argument for a formula, there are two interesting types of argumentation against it;

Definition 5 *An argument* $(\phi_\mathtt{i}, \mathtt{P}_\mathtt{i})$ rebuts *an argument* $(\phi_\mathtt{j}, \mathtt{P}_\mathtt{j})$ *if* $\phi_\mathtt{i}$ *attacks* $\phi_\mathtt{j}$.

The notion of "attack" in a normative system is defined in the next subsection.

Definition 6 *An argument* $(\phi_\mathtt{i}, \mathtt{P}_\mathtt{i})$ undercuts *an argument* $(\phi_\mathtt{j}, \mathtt{P}_\mathtt{j})$ *if there exists* $\mathtt{s}_\mathtt{k} \in \mathtt{P}_\mathtt{j}$ *such that (1)* $\mathtt{s}_\mathtt{k}$ *is a formula and* $\phi_\mathtt{i}$ *attacks* $\mathtt{s}_\mathtt{k}$*, or (2)* $\mathtt{s}_\mathtt{k} = \Gamma_\mathtt{k} \vdash_{\mathtt{d}_\mathtt{k}} \psi$ *and* $\phi_\mathtt{k}$ *attacks* ψ*.*

In order of increasing acceptability:

A1 The class of all arguments that may be made from Γ.

A2 The class of all non-trivial arguments that may be made from Γ.

A3 The class of all arguments that may be made from Γ for propositions for which there are no rebutting arguments that may be made from Γ.

A4 The class of all arguments that may be made from Γ for propositions for which there are no undercutting arguments that may be made from Γ.

A5 The class of all tautological arguments that may be made from Γ.

Informally, the idea is that arguments in higher numbered classes are more acceptable because they are less questionable.

6.1. *Argumentation in normative systems*

Parsons et al. use this argumentation framework in non-regulated environments. This approach has two undesired consequences,

- The argument of power: if there is a power relationship between the agents (Alonso 1998, 1999; Castelfranchi et al. 1992), that is, if one of the agents x needs y whilst y just prefers to interact with x, then, the argument put forward by the agent with more authority has the winning number. That means that the content and essence of the argumentative machine does not really matter after all.
- The power of the argument?: In symmetric situations, the agents are assumed to be benevolent and trustful. It is, however, well known that this assumption only works in Cooperative Problem Solving – CPS domains (Wooldridge and Jennings 1994) in which the agents have a common goal and try to maximize the global utility.

To overcome these limitations we propose to adapt Parsons's et al. model to open normative domains.[5] On the one hand, argumentation can help to solve problems in normative systems. On the other hand, if we want the agents to accept reasonable arguments regardless of their relations, then we need norms.

Definition 7 *Given agents* $\mathtt{x}$ *and* $\mathtt{y}$*, we say that a formula* $\phi_{\mathtt{i}}$ *of the language of agent* $\mathtt{x}$ *attacks a formula* $\phi_{\mathtt{j}}$ *of the language of* $\mathtt{y}$ *if the following case holds:* $\phi_{\mathtt{i}} = \mathtt{R}(\mathtt{x}, \alpha)$ *and* $\phi_{\mathtt{j}} = \mathtt{R}(\mathtt{y}, \beta)$*, and* $[\alpha > \beta \wedge \mathtt{Inh}(\beta, \alpha) \wedge \mathtt{I}(\mathtt{x}, \alpha) \wedge \mathtt{I}(\mathtt{y}, \beta)]$.

In Parsons's et al. model, a formula attacks another when they generate conflict intentions. In our proposal, this requisite is necessary but not sufficient: Besides this condition, $\phi_{\mathtt{i}}$ attacks $\phi_{\mathtt{j}}$ when the right described by the first formula has priority over the second one. That means that whereas a free agent has to find a counterargument (with threats or alternative reasons) to stop the other having an undesirable intention, in our model the second agent must understand that it is forbidden to realise such an intention.

6.2. *Parking example*

Let's illustrate this argumentative model with an example taken from the traffic domain (Figure 4). Imagine two agents x and y who intend to park in the same place, one forwards, the other backwards. They have conflict intentions and goals and would try to sort this situation out by arguing.

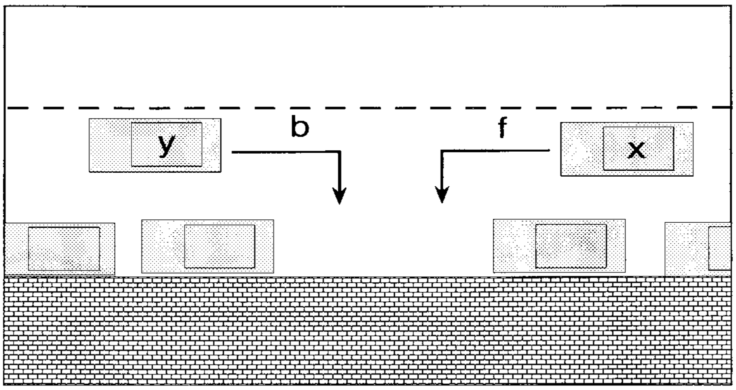

Figure 4. The parking example. Agent *x* is driving forwards (f) as it parks its car in the gap. Agent *y* intends to park its car in the same gap driving backwards (b). As they have conflicting goals and cannot be referred to a specific norm, agents *x* and *y* will initiate an argumentation episode.

First we need to extend $\mathcal{L}$ by introducing some simple facts about planning and driving. That an agent x has a goal α is represented as $\mathtt{G}(\mathtt{x}, \alpha)$. Goal/subgoal relations are conceived as AND/OR trees. Each goal has an associated tree, which is formed probably by different OR-subtrees or path-solutions. Trees describe different levels of goal abstraction, and leaves are executable actions that form alternative plans when they appear AND related.

- **Theory of planning:**

  If an agent is allowed to execute an action and has the goal to bring this state of the world, then it will have the intention to execute that action.

  $$\mathtt{A}(\mathtt{x}, \alpha) \wedge \mathtt{G}(\mathtt{x}, \alpha) \rightarrow \mathtt{I}(\mathtt{x}, \alpha) \tag{5}$$

  If an agent has a goal α and there are different disjunctive subgoals associated to it, $\mathtt{sub}(\alpha, \{\alpha_1 \vee \ldots \vee \alpha_n\})$, then the agent will eventually have to adopt one of these subgoals.

  $$\mathtt{G}(\mathtt{x}, \alpha) \wedge \mathtt{sub}(\alpha, \{\alpha_1 \vee \ldots \vee \alpha_n\}) \rightarrow \mathtt{G}(\mathtt{x}, \alpha_1) \vee \ldots \vee \mathtt{G}(\mathtt{x}, \alpha_n) \tag{6}$$

  If an agent has the intention of executing a plan (a complex action), $\alpha = \{\alpha_1 \wedge \ldots \wedge \alpha_n\}$, then he has the intention of executing all the actions of that plan.

  $$\forall \alpha_i \in \alpha, \mathtt{I}(\mathtt{x}, \alpha) \rightarrow \mathtt{I}(\mathtt{x}, \alpha_i) \tag{7}$$

  We can add the same axiom for rights: If an agent has the right of executing a complex action he has also the right to execute its parts.

  $$\forall \alpha_i \in \alpha, \mathtt{R}(\mathtt{x}, \alpha) \rightarrow \mathtt{R}(\mathtt{x}, \alpha_i) \tag{8}$$

On the other hand, if an agent has a right to execute a disjunctive action, then he has the right to execute any of the disjunctors.

$$\mathtt{R}(\mathtt{x}, \alpha) \wedge \mathtt{sub}(\alpha, \{\alpha_1 \vee, \ldots, \vee \alpha_n\}) \rightarrow \mathtt{R}(\mathtt{x}, \alpha_1) \wedge \ldots \wedge \mathtt{R}(\mathtt{x}, \alpha_n) \qquad (9)$$

Finally, two actions are incompatible if they inhibit each other:

$$\mathtt{Inc}(\alpha, \beta) \equiv \mathtt{Inh}(\alpha, \beta) \wedge \mathtt{Inh}(\beta, \alpha) \qquad (10)$$

- **Theory of driving:**
First, to park forwards, $\mathtt{p_f}$, or backwards, $\mathtt{p_b}$, are OR-subgoals of Parking, $\mathtt{p}$.

$$\mathtt{sub}(\mathtt{p}, (\mathtt{p_f} \vee \mathtt{p_b})) \qquad (11)$$

Second, driving is part of parking. Consequently, driving in a particular direction, $\mathtt{d_i}$, is part of parking in that direction.

$$\mathtt{d_i} \in \mathtt{p_i} \qquad (12)$$

As a consequence, following (8)

$$\mathtt{R}(\mathtt{x}, \mathtt{p_i}) \rightarrow \mathtt{R}(\mathtt{x}, \mathtt{d_i}) \qquad (13)$$

And, following (7)

$$\mathtt{I}(\mathtt{x}, \mathtt{p_i}) \rightarrow \mathtt{I}(\mathtt{x}, \mathtt{d_i}) \qquad (14)$$

We also assume that the traffic code says that driving forwards has priority over driving backwards,

$$\mathtt{d_f} > \mathtt{d_b} \qquad (15)$$

Finally, driving backwards and driving forwards are incompatible actions.

$$\mathtt{Inc}(\mathtt{d_i}, \mathtt{d_j}) \qquad (16)$$

Now, the argumentation goes as follows. First, we describe *y*'s argument. Then, *x*'s counterargument is presented.

- *y*'s argument:

$$\mathtt{G}(\mathtt{y}, \mathtt{p}) \qquad (17)$$

$$\mathtt{G}(\mathtt{y}, \mathtt{p}) \wedge \mathtt{sub}(\mathtt{p}, (\mathtt{p_f} \vee \mathtt{p_b})) \quad \text{introduction of } \wedge \text{ (17)(11)} \qquad (18)$$

$$\mathtt{G}(\mathtt{y}, \mathtt{p_b}) \quad \text{mp (18)(6) and elimination of } \vee \qquad (19)$$

$$\mathtt{R}(\mathtt{y}, \mathtt{p}) \qquad (20)$$

$$\mathtt{R(y,p)} \wedge \mathtt{sub(p,(p_f} \vee \mathtt{p_b))} \qquad \text{introduction of } \wedge \text{ (20)(11)} \tag{21}$$

$$\mathtt{R(y,p_b)} \qquad \text{mp (21)(9) and elimination of } \wedge \tag{22}$$

$$\mathtt{A(y,p_b)} \tag{23}$$

$$\mathtt{G(y,p_b)} \wedge \mathtt{A(y,p_b)} \qquad \text{introduction of } \wedge \text{ (19)(23)} \tag{24}$$

$$\mathtt{I(y,p_b)} \qquad \text{mp (5)(24)} \tag{25}$$

Then *y*'s argument is $(\phi_{\mathtt{j}}, \mathtt{Pj}) = (\mathtt{I(y,p_b)}, \{(17)–(25)\})$.
It is easy to see from (1) that (23) holds, as $\mathtt{p_b}$ and $\mathtt{p_f}$ are, by default, equally rated. However, as *x* has also that intention (*x* will build the same parallel argument (17')–(25') about his intention to park forwards), both arguments will be in class A2 (they rebut each other but they are consistent). Consequently, unless another normative argument is found, the agents will get stuck. *x* will find such an argument.

- *x*'s argument:

$$\mathtt{I(x,d_f)} \qquad \text{mp (14)(25')} \tag{26}$$

$$\mathtt{R(x,d_f)} \qquad \text{mp (13)(22')} \tag{27}$$

x's simple argument, $(\phi_{\mathtt{i}}, \mathtt{P_i}) = (\mathtt{R(x,d_f)}, \{(17')–(25'), (26), (27)\})$, undercuts $(\phi_{\mathtt{j}}, \mathtt{Pj})$ because $\phi_{\mathtt{i}} = \mathtt{R(x,d_f)}$ attacks a formula $\psi = \mathtt{R(y,d_b)}$ derivable from (22) (just applying substitution and modus ponens to (13) and (22)). Notice that all the conditions in **Definition 7**, namely, $\mathtt{d_f} > \mathtt{d_b}$ (15), $\mathtt{I(x,d_f)}$ (27), $\mathtt{I(y,d_b)}$ (from (25), (7), (12)), and $\mathtt{Inh(d_b,d_f)}$ (from (10) and (16)) hold.

So, *x* comes up with this solution: It can derive a formula from its opponent's argument (*y*'s right to drive backwards from the right to park backwards) that is attacked by its right to drive forwards. This argument is A4, since *y* can neither rebut nor undercut it. *y* is assumed to accept *x*'s proposal and respect this result because the argumentation process is supervised by the group. However, if *y* does not accept the argumentation outcome and persists in parking backwards, *x* will request for help.

7. Conclusions and Further Work

In this paper, we have introduced a preliminary study of the concept of right. There have been other approaches to rights in the MAS literature: (Castelfranchi 1995) explained how social commitments generate rights (to claim), and (Norman et al. 1998) have presented rights (permissions) as arguments in agreements. We have focused our work on more basic rights, called liberties, which represent and protect universal interests. No doubt all these rights are closely related. However, we understand that liberties play a very special role in a theory of coordination and social action. They endow autonomous agents with enough freedom and groups with enough power to assure stable and efficient solutions in uncertain domains. In the second part of the paper we have also applied Parsons's et al. model of argumentation (Parsons et al. 1998a, b; Sierra et al. 1998) to open normative systems to minimize the problems caused by underspecification or by the agents' limited knowledge about legal systems.

The most obvious issue to be addressed in future work is the refinement of the model, both theoretically and formally. Specifically, some discussion must be undertaken about the dynamic aspects of rights (already treated in (Conte et al. 1998)): Their genesis, acceptance and abandonment.

Besides, it is obvious that the relationships between normative and social attitudes (joint intentions, mutual beliefs, etc.) must be studied in depth.

Finally, even though the traffic-world is quite natural to explain and to understand the intuitive meaning of rights, it is also true that a more elaborated domain problem (for example, electronic commerce) will show better how useful the framework here described may be.

Notes

1. We assume that *x* has priority according to the British current traffic code.
2. This idea finds philosophical support in the Theory of Social Exchange, according to which every social interaction is rooted in the "reciprocity principle" (Blau 1968; Simmel 1908).
3. It is also true that communication can be required even if rights are fully specified. Typically, when one's rights affect others'. For example, if *x* decides to park just before the junction, it will have to communicate its intention to *y*, as this second agent can then exercise its right of driving from *B* to *A* without delay.
4. In (Alonso 2002) we identified other cases in which normative agents have to negotiate, for example, when they exercise "social rights". Exercising our rights does not necessarily result in achieving our goals since others' cooperation may be required. Benevolence is not assumed, so that one's right does not automatically trigger others' actions. Leaving aside physical capacity, this is specially true in the case of "social rights". We can use Levesque's et al. "Convoy example" (Levesque et al. 1990) to illustrate this concept: An agent *x* can have the right to drive together with another agent *y* as a convoy. However, *x* needs *y* to exercise his right to drive, and to do it in a coordinated way. That means that *x* will likely have to convince *y* to adopt that goal, and then they will have to arrange how to drive the convoy (who drives in front, when to stop, etc. . . .).
5. Of course, in closed normative systems, such as FishMarket (Noriega and Sierra 1997), there is no space for negotiation or argumentation. The dialogical framework is well-defined and known by every agent.
6. This definition can be easily extended to one-shot deals, so that if the group believes that an agent is not abiding by an agreement, the group is obliged to force him to fulfil his commitments.

References

Alonso, E. (1998). How Individuals Negotiate Societies. In Proceedings of *The Third International Conference on Multi-Agent Systems, ICMAS-98*, 18–25. Los Alamitos, CA: IEEE Computer Society.

Alonso, E. (1999). An Individualistic Approach to Social Action in Multi-Agent Systems. *Journal of Experimental and Theoretical Artificial Intelligence* **11**: 519–530.

Alonso, E. (2002). Rights for Multi-agent Systems. In d'Inverno, M., Luck, M., Fisher, M. & Preist, C. (eds.) *Foundations and Applications of Multi-Agent Systems: UKMAS Workshops 1996–2000 Selected Papers*, 59–72, LNAI 2403, Springer: Berlin.

Barry, N. P. (1989). *An Introduction to Modern Political Theory*. Macmillan: London.

Blau, P. M. (1968). Interaction: Social Exchange. In Sills, D. L. (ed.) *The International Encyclopedia of the Social Sciences*. Macmillan: New York.

Castelfranchi, C., Miceli, M. & Cesta, A. (1992). Dependence Relations among Autonomous Agents. In Werner, E. & Demazeau, Y. (eds.) *Decentralized A.I. 3*, 215–227. North-Holland: Amsterdam, The Netherlands.

Castelfranchi, C. (1995). Commitments: ¿From Individual Intentions to Groups and Organizations. In Proceedings of *The First International Conference on Multi-Agent Systems, ICMAS-95*, 41–48. Cambridge, MA: MIT Press.

Conte, R., Castelfranchi, C. & Dignum, F. (1998). Autonomous Norm-acceptance. In Muller, J. P., Singh, M. P. & Rao, A. (eds.) *Intelligent Agents V: Agent Theories, Architectures, and Languages: 5th International Workshop ATAL-98*, 319–333. Berlin: Springer.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press.

Jennings, N. R. (1993). Commitments and Conventions: The Foundation of Coordination in Multi-Agent Systems. *The Knowledge Engineering Review* **8**: 223–250.

Kraus, S., Wilkenfeld, J. & Zlotkin, G. (1995). Multiagent Negotiation under Time Constraints. *Artificial Intelligence* **75**: 297–345.

Kraus, S., Sycara, K. & Evenchik, A. (1998). Reaching Agreements through Argumentation: A Logical Model and Implementation. *Artificial Intelligence* **104**: 1–69.

Krogh, C. (1996). The Rights of Agents. In Wooldridge, M. J., Muller, J. P. & Tambe, M. (eds.) *Intelligent Agents II: Agent Theories, Architectures, and Languages: IJCAI-95 Workshop*, 1–16. Berlin: Springer.

Levesque, H. J., Cohe, P. R. & Nunes, H. T. (1990). On Acting Together. In Dietterich, T. & Swartout, W. (eds.) *Proceedings of the 8th National Conference on Artificial Intelligence, AAAI-90*, 94–99. Cambridge, MA: MIT Press.

Nielsen, K. & Shiner, R. A. (1977). *New Essays on Contract Theory*. Canadian Association for Publishing in Philosophy.

Noriega, P. & Sierra, C. (1997). Towards Layered Dialogical Agents. In Proceedings of *The 12th European Conference on Artificial Intelligence, ECAI-96*, 173–188. John Wiley & Sons: New York.

Norman, T. J., Sierra, C. & Jennings, N. R. (1998). Rights and Commitment in Multi-agent Agreements. In Proceedings of *The Third International Conference on Multi-Agent Systems, ICMAS-98*, 222–229. Los Alamitos, CA: IEEE Computer Society.

O'Hare, G. M. P. & Jennings, N. R. (1996). *Foundations of Distributed Artificial Intelligence*. John Wiley & Sons: New York.

Parsons, C., Sierra, C. & Jennings, N. R. (1998a). Agents that Reason and Negotiate by Arguing. *Journal of Logic and Computation* **8**: 261–292.

Parsons, C., Sierra, C. & Jennings, N. R. (1998b). Multi-context Argumentative Agents. In Proceedings of *CommonSense-98*, 298–349.

Reiner, R. (1995). Arguments Against the Possibility of Perfect Rationality. *Minds and Machines* **5**: 373–389.

Rosenschein, J. S. & Zlotkin, G. (1994). *Rules of Encounter: Designing Conventions for Automated Negotiation Among Computers*. MIT Press: Cambridge, MA.

Shoham, Y. & Tennenhlotz, M. (1992). On the Synthesis of Useful Social Laws for Artificial Agents Societies. In Proceedings of *The Tenth National Conference on Artificial Intelligence, AAAI-92*, 276–281. Menlo Park, CA: AAAI Press.

Sichman, J. S., Conte, R., Demazeau, Y. & Castelfranchi, C. (1994). A Social Reasoning Mechanism Based on Dependence Networks. In Wooldridge, M. J. & Jennings N. R. (eds.) *Intelligent Agents: ECAI-94 Workshop on Agent Theories, Architectures and Languages*, 173–177. Berlin: Springer.

Sierra, C., Jennings, N. R., Noriega, P. & Parsons, S. (1988). A Framework for Argumentation-based Negotiation. In Singh, M. P., Rao, A. & Wooldridge, M. J. (eds.) *Intelligent Agents IV: 4th International Workshop on Agent Theories, Architectures and Languages, ATAL-97*, 177–192. Berlin: Springer.

Simmel, G. (1908). *The Sociology of Georg Simmel*. Free Press: New York.

Staniford, G. (1994). Multi-agent System Design: Using Human Societal Metaphors and Normative Logic. In Wooldridge, M. J. & Jennings N. R. (eds.) *Intelligent Agents: ECAI-94 Workshop on Agent Theories, Architectures and Languages*, 289–293. Berlin: Springer.

Sycara, K. (1998). Multiagent Systems. *AI Magazine* **19**: 79–92.

Tennenhlotz, M. (1998). On Stable Social Laws and Qualitative Equilibria. *Artificial Intelligence* **102**: 1–20.

Walker, A. & Wooldridge, M. (1995) Understanding the Emergence of Conventions in Multi-agent Systems. In Proceedings of *The First International Conference on Multi-Agent Systems, ICMAS-95*, 384–389. Cambridge, MA: MIT Press.

Weiss, G. (1999). *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. The MIT Press: Cambridge, MA.

Wooldridge, M. & Jennings, N. R. (1994). Towards a Theory of Cooperative Problem Solving. In Perram, J. W. & Muller, J. P. (eds.) *Distributed Software Agents and Applications: 6th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW-94*, 40–53. Berlin: Springer.