Negotiation and Joint Commitments in Multi-Agent Systems

Pietro Panzarasa and Nicholas R. Jennings

Department of Electronics and Computer Science, University of Southampton
Southampton SO17 1BJ, U.K.
{pp,nrj}@ecs.soton.ac.uk

Abstract. In this paper, it is argued that negotiation can be regarded as a socio-cognitive process for the transformation of joint commitments in multi-agent environments. To this end, a quantified multi-modal logical language is developed for reasoning about and representing agents' mental attitudes. Drawing on this language, negotiation is formalised using the classical axiomatic-deductive methodology for theory building. Assumptions are presented, and properties discussed on a proof-theoretic basis. The explanatory breadth of the formalism is illustrated by looking at its applicability in situations in which agents are boundedly rational, have asymmetric and incomplete information, are motivated by conflicting interests, and behave opportunistically.

1 Introduction

Recent advances in distributed artificial intelligence (DAI), social networks, cognitive sciences, and organisation theory have led to a new perspective on multi-agent systems (MASs) that takes into account both their computational nature and the underlying social and knowledge networks [7, 12, 22, 35]. The hallmark of this perspective is the idea that cognition occurs at multiple levels, not only within the individual agent, but also as an emergent phenomenon from the interaction among multiple agents. Drawing on this view, in this paper we develop a theory of negotiation in a multi-agent setting, where mechanisms of social behaviour are predicated on a fully explicated model of the agents' cognition at both the individual and the joint level. At the heart of our theory lies the conceptualisation of sociality in terms of higher-order joint mental attitudes and behaviours that rest on and transcend the individual mental attitudes and behaviours of cognitively and socially interconnected agents. Building on this conceptual framework, this paper will show how it is possible to take some steps towards a new account of the *cognitive foundations of pluralism* in MASs [6, 34]. This is the main theoretical contribution of our work, besides its aim of presenting a new conception of negotiation within a multi-agent environment.

Our approach is to conceptualise basic principles governing intelligent communities of artificial agents who negotiate to achieve a common objective. Such principles provide specifications for the design of artificial agents, and approximate a theory of human negotiation. As a result, to the extent that it provides assistance to

practitioners interested in building distributed agent architectures, this paper is mainly intended to contribute to DAI research. However, in its emphasis on exploring and formally specifying a number of important cognitive and behavioural properties of social agenthood, our work also contributes to organisation theory and cognitive science.

The remainder of the paper is organised as follows. In section 2 a quantified multi-modal logical language will be presented that draws on and extends standard Belief-Desire-Intention (BDI) logics [11, 31, 35]. Drawing on this language, negotiation will be formalised, using the classical axiomatic-deductive methodology for theory building [27]. In section 3 a number of assumptions about agents' behavioural and cognitive processes during negotiation will be presented, and in section 4 some properties of the formalism will be examined on a proof-theoretic basis. Finally, very much in the spirit advocated by a number of scholars in computational organisation science [5, 7], in section 5 an attempt will be made to summarise and discuss our major findings in terms of their contributions towards an interdisciplinary integration of methods, principles and research questions from differing disciplines, such as computer science, DAI, organisation theory, economics, and sociology.

2 The Logical Language

This section gives an overview of the formal framework that will be used throughout the remainder of the paper. The logic is a simplified version of what we developed in [23], to which the reader should refer for a complete formal definition. The formalism used is a first-order, linear-time, quantified, many-sorted, multi-modal logic for reasoning about agents, groups, actions, and mental attitudes, with explicit reference to time points and intervals.

First, a brief description of the model of time that underpins our logic. Every occurrence of a formula ϕ is stamped with a time t_i, written $\phi(t_i)$, meaning that ϕ holds at time t_i[1]. Time is taken to be composed of points and, for simplicity, is assumed to be discrete and linear [3]. For time point t_i, $t_i + 1$ is the time point that increments t_i; that is, $t_i + 1$ is the time point obtained by extending t_i by a time point. Temporal intervals are defined as pairs of points. Thus, for example, $\phi(6, 8)$ means that ϕ is satisfied at time points 6, 7, and 8. Furthermore, intervals of the form (t_i, t_i) can equally be written as time points.

The logic is many-sorted: terms come in six sorts. First, we have terms that denote *agents*, and we use a_i, a_j, … as variables ranging over individual agents. Second, we have terms that denote *groups* of agents, and we use gr_i, gr_j, … as variables ranging over such groups. A group of agents is simply a non-empty subset of the set of agents. Third, we have terms that denote *time points*, and we use t_i, t_j, … and so on as variables ranging over time points. Fourth, we have terms that denote *temporal intervals*, defined as pairs of time points, and we use i_i, i_j, … as variables ranging over time intervals. Fifth, we have terms that denote *actions*, and we use e_i, e_j, … as variables ranging over actions. Finally, we have terms that denote *generic objects* in

[1] We assume that a missing temporal term in a well-formed formula is the same as the closest temporal term to its right.

the environment (e.g. tables, blocks, and so on), and we use o_i, o_j, … as variables ranging over objects.

The language includes first-order equality: a formula ($\tau = \tau'$) will be true if τ and τ' denote the same individual. The operators $\neg$ (not) and $\vee$ (or) have classical semantics, as does the universal quantifier $\forall$. The remaining classical connectives and existential quantifier are assumed to be introduced as abbreviations, in the obvious way. We also use the punctuation symbols ")", "(", "[", "]", and comma ",". With the "$\in$" operator, we relate agents to groups: $a_i \in gr_i$ means that the agent denoted by a_i is a member of the group denoted by gr_i. The operator $Singleton(gr_i, a_i)(t_i)$ means that gr_i at t_i is a singleton group with a_i as the only member. For arbitrary a_i, e_i, t_i, we have: $Singleton(gr_i, a_i)(t_i) \equiv \forall a_j \ (a_j \in gr_i)(t_i) \supset (a_j = a_i)(t_i)$.

To express the occurrence of an action in the world, our language includes the operator $Occurs(e_i)(t_i)$, which means that action e_i happens at time t_i. Furthermore, we have action expressions that depend on the truth or falsity of formulae: $Occurs(\phi?)(t_i)$ means that ϕ is satisfied at t_i. Actions may be performed by an individual agent (single-agent actions) or by a group of agents (multi-agent actions). For simplicity, we assume that an action is either a single-agent or a multi-agent action, but not both. A sentence of the form $Agts(gr_i, e_i)(t_i)$ states that at t_i the group denoted by gr_i are the agents required to perform the action denoted by e_i. To express that agent a_i is the only agent of e_i at t_i, we have (for arbitrary a_i, e_i, t_i): $Agt(a_i, e_i)(t_i) \equiv \forall gr_i \ Agts(gr_i, e_i)(t_i) \supset Singleton(gr_i, a_i)(t_i)$.

To capture the notion of a state-directed action, we now introduce the derived operator $plan(gr_i, e_i, \phi(t_j))(t_i)$, which expresses the fact that, at time t_i, action e_i represents, for group gr_i, a plan for making ϕ true at t_j ($t_j > t_i$) [19, 23, 36]. For arbitrary gr_i, e_i, t_i and t_j ($t_j > t_i$), we have: $plan(gr_i, e_i, \phi(t_j))(t_i) \equiv \exists t_h, t_k \ (t_i \leq t_h \leq t_k < t_j)$ s.t. $Occurs(e_i)(t_h, t_k) \wedge Agts(gr_i, e_i)(t_h, t_k) \wedge [Occurs(e_i)(t_h, t_k) \supset Occurs(\phi?)(t_j)]$. Informally, we say that at time t_i action e_i is a plan for group gr_i to make ϕ true at $t_j > t_i$ iff: (a) e_i will occur sometime before t_j; (b) gr_i is the group required to perform e_i; and (c) if e_i occurs, then ϕ will be satisfied afterwards at t_j [2]. Finally, for arbitrary a_i, e_i, t_i and t_j ($t_j > t_i$), we have: $plan(a_i, e_i, \phi(t_j))(t_i) \equiv \forall gr_i \ plan(gr_i, e_i, \phi(t_j))(t_i) \supset Singleton(gr_i, a_i)(t_i)$. The above definitions of single-agent and multi-agent plans capture the notions of actions that agents or groups *eventually* perform to satisfy certain states of the world. We now want to express the *past* execution of state-directed actions. To this end, we introduce the operator $<plan(gr_i, e_i, \phi)>(t_i)$, which means that at t_i state ϕ has been brought about as a result of the performance of action e_i by group gr_i. For arbitrary gr_i, e_i, and t_i, we have: $<plan(gr_i, e_i, \phi)>(t_i) \equiv \exists t_j, t_h \ (t_j \leq t_h < t_i)$ s.t. $Occurs(e_i)(t_j, t_h) \wedge Agts(gr_i, e_i)(t_j, t_h) \wedge [Occurs(e_i)(t_j, t_h) \supset Occurs(\phi?)(t_i)]$. Informally, we say that, at time t_i, ϕ has been made true as a consequence of the performance of action e_i by group gr_i iff: (a) e_i occurred sometime in the past; (b) gr_i was the group required to perform e_i; and (c) ϕ was satisfied afterwards at t_i as a consequence of the performance of e_i. Finally, for arbitrary a_i, e_i and t_i, we have: $<plan(a_i, e_i, \phi)>(t_i) \equiv \forall gr_i <plan(gr_i, e_i, \phi)(t_i)> \supset Singleton\ (gr_i, a_i)(t_i)$.

---

[2] For simplicity, we do not adopt a more sophisticated definition of plans (e.g. partial or hierarchical non-linear plans [19]). We leave such refinements to future work.

The logic is further enriched by a set of modal operators for reasoning about agents' mental attitudes. Drawing on a fairly standard BDI framework [11, 29, 31, 35], we introduce the operators $Bel(a_i, \phi)(t_i)$ and $Int(a_i, \phi)(t_i)$, which mean that at time t_i agent a_i has, respectively, a belief that ϕ holds and an intention towards ϕ, where ϕ is a well-formed formula. Beliefs may concern facts of the world and can be nested, namely they can be introspective or have others' mental attitudes as their arguments. Also, beliefs may be incomplete. The formal semantics for beliefs are a natural extension of the traditional Hintikka's possible-worlds semantics [18]. The restrictions imposed on the belief-accessibility relation ensure a belief axiomatisation of KD45 (corresponding to a "Weak S5 modal logic"), which thus implies that beliefs are consistent and closed under consequence, and that agents are aware of what they do and do not believe [9]. Finally, an agent's intention represents the states of the world that the agent is "self-committed" to achieving or maintaining [0]. Like beliefs, intentions can be nested, and their semantics are given in terms of possible worlds. Restrictions on the intention-accessibility relation ensure that the logic of intentions validates axioms K and D, which thus implies that intentions are consistent and closed under consequence. Finally, we introduce a weak realism constraint ensuring that agents' intentions do not contradict their beliefs [29].

In addition to beliefs and intentions, agents have local preferences. The operator $Pref(a_i, \phi, \psi)(t_i)$ means that at time t_i agent a_i prefers ϕ over ψ, where ϕ and ψ are well-formed formulae. Preferences can be nested. The semantics for preferences are given in terms of closest worlds (see [4] and [23] for details). Our language also contains the operator $Comm(a_i, gr_i, e_i)(t_i)$, which means that at time t_i agent a_i is committed towards group gr_i to performing action e_i [8]. Building on this, we introduce the derived operator $Comm(a_i, gr_i, \phi(t_j))(t_i)$ to express the commitment that agent a_i has towards group gr_i to making ϕ true at t_j $(t_j>t_i)$ [15, 23]. For arbitrary a_i, gr_i, t_i and t_j $(t_j>t_i)$, we have: $Comm(a_i, gr_i, \phi(t_j))(t_i) \equiv \exists\, e_i$ s.t. $[Comm(a_i, gr_i, e_i)(t_i) \wedge (plan(a_i, e_i, \phi(t_j))(t_i) \vee \exists\, e_j, \exists\, t_k\ (t_i < t_k < t_j)$ s.t. $(plan(a_i, e_i, plan(gr_i, e_j, \phi(t_j))(t_k))(t_i) \vee plan(a_i, e_i, plan(\{gr_i, a_i\}, e_j, \phi(t_j))(t_k))(t_i)))]$. Informally, we say that at t_i agent a_i is socially committed towards group gr_i to making ϕ true at $t_j>t_i$ iff there is at least one action e_i such that at t_i: (i) a_i is committed towards gr_i to performing e_i; and (ii) either e_i is a plan for a_i to achieve ϕ at t_j; or (iii) e_i is a plan for a_i to allow gr_i to achieve ϕ at t_j; or (iv) e_i is a plan for a_i to allow gr_i and a_i to achieve ϕ collaboratively at t_j.

In addition to individual agents' attitudes, we now introduce joint mental attitudes. First, our language includes the modal operator $M\text{-}BEL\ (gr_i, \phi)(t_i)$, which means that, at time t_i, group gr_i has a *mutual belief* that ϕ holds. Crudely, a mutual belief can be defined as an infinite conjunction of an agent's belief about an agent's belief about an agent's belief and so forth, that a proposition holds [12]. To define the semantics for mutual beliefs, we introduce the operator $E\text{-}BEL\ (gr_i, \phi)\ (t_i)$, which means that, at time t_i, every agent in gr_i believes that ϕ holds. For arbitrary gr_i and t_i, we have: $E\text{-}BEL\ (gr_i, \phi)(t_i) \equiv \forall a_i \in gr_i\ Bel\ (a_i, \phi)(t_i)$. If $k \in N$ such that $k > 0$, we define $E\text{-}BEL^k(gr_i, \phi)(t_i)$ inductively in the following way. Let $E\text{-}BEL^k(gr_i, \phi)(t_i)$ be an abbreviation for $E\text{-}BEL(gr_i, \phi)(t_i)$ if $k = 1$, and for $E\text{-}BEL\ (gr_i, E\text{-}BEL^{k-1}(gr_i, \phi))(t_i)$ otherwise. Thus, we define mutual beliefs as follows: $M\text{-}BEL\ (gr_i, \phi)(t_i) \equiv \wedge_{k>0}\, E\text{-}BEL^k\ (gr_i, \phi)(t_i)$ [35, 36].

Furthermore, to express the notion of *joint intentions*, we introduce the operator $J\text{-}INT(gr_i, \phi)(t_i)$, which means that at time t_i group gr_i holds a joint intention towards ϕ.

Informally, we say that a group has a joint intention towards ϕ iff: (a) it is true (and mutual belief in gr_i) that each member has the intention towards ϕ; and (b) it is true (and mutual belief in gr_i) that each member intends that the other members have an intention towards ϕ. Formally, for arbitrary gr_i and t_i, we have: $J\text{-}INT(gr_i, \phi)(t_i) \equiv E\text{-}INT(gr_i, \phi)(t_i) \wedge M\text{-}BEL(gr_i, E\text{-}INT(gr_i, \phi))(t_i) \wedge E\text{-}INT(gr_i, E\text{-}INT(gr_i, \phi))(t_i) \wedge M\text{-}BEL(gr_i, E\text{-}INT(gr_i, E\text{-}INT(gr_i, \phi))))(t_i)$, where $E\text{-}INT\ (gr_i, \phi)(t_i) \equiv \forall a_i \in gr_i\ Int\ (a_i, \phi)(t_i)$.

Finally, we can give a formalisation of the notion of *joint commitment* [8, 12, 15, 23]. We say that, at time t_i, a group gr_i has a joint commitment to making ϕ true at t_j ($t_j > t_i$) iff: (i) in gr_i it is mutually believed that ϕ will be true at t_j; (ii) gr_i has the joint intention that ϕ will be true at t_j; (iii) it is true (and mutual belief in gr_i) that each member of gr_i is socially committed towards gr_i to making ϕ true at t_j; and (iv) it is true (and mutual belief in gr_i) that (ii) will continue to hold until it is mutually believed in gr_i either that ϕ will not be true at t_j, or that at least one of the members drops its commitment towards gr_i to making ϕ true at t_j. Formally, for arbitrary gr_i, t_i and t_j ($t_j > t_i$), we have:

$$J\text{-}COMM(gr_i, \phi(t_j))(t_i) \equiv M\text{-}BEL(gr_i, \phi(t_j))(t_i) \wedge J\text{-}INT(gr_i, \phi(t_j))(t_i) \wedge \forall a_i \in gr_i\ [Comm(a_i, gr_i, \phi(t_j)) \wedge M\text{-}BEL(gr_i, Comm(a_i, gr_i, \phi(t_j)))](t_i) \wedge \gamma(t_i) \wedge M\text{-}BEL(gr_i, \gamma)\ (t_i),$$

where $\gamma \equiv [J\text{-}INT(gr_i, \phi(t_j))(t_i, t_j) \vee \exists t_k\ (t_i < t_k \leq t_j)$ s.t. $((M\text{-}BEL(gr_i, \neg\phi(t_j)) \vee \exists a_i \in gr_i$ s.t. $(\neg Comm(a_i, gr_i, \phi(t_j)) \wedge M\text{-}BEL(gr_i, \neg Comm(a_i, gr_i, \phi(t_j)))))(t_k) \wedge \forall t_h\ (t_i \leq t_h < t_k)$ $J\text{-}INT(gr_i, \phi(t_j))\ (t_h))]$.

3 A Model of Negotiation

Building on the language outlined above, we will now formalise a number of assumptions that are intended to model the agents' cognition and social behaviour during negotiation [23, 34]. Negotiation is here conceived of as a *socio-cognitive process for the transformation of commitments* in a social setting comprising cognitive agents [23]. This conception can be articulated into two core ideas. First, negotiation is seen as grounded on a joint commitment among the members of a group to achieving a state of the world[3]. Second, negotiation is regarded as primarily aimed at generating a joint commitment among the agents to acting according to a joint plan of action. Thus, motivated by a prior joint commitment, the agents are moved into negotiation. In turn, negotiation transforms the agents' prior joint commitment towards a state into a derived joint commitment towards a plan that is a means for achieving that state. For example, in an economic transaction, both the buyer and the seller are moved into negotiation by a prior joint commitment to having a good (or service) delivered from the seller to the buyer, and the counter-value from

[3] How this commitment is generated is not our concern in this paper.

the buyer to the seller. If successful, negotiation will transform the prior joint commitment into a new conclusive commitment leading the buyer and the seller to perform a plan that specifies the conditions (e.g. price, time, place, quality, guarantees) at which the economic transaction is to be finalised.

As it stands, this perspective is consistent with the idea that negotiation is inherently intertwined with the process of practical reasoning typically undertaken in social settings [2, 23]. In fact, the transformation of commitments that negotiation brings about can be seen as instrumental to the collaborative search for a solution to a common practical problem. The common problem concerns what is to be done by a group to fulfil a prior joint commitment towards an end; the agreed solution reflects a conclusive joint commitment towards the means to secure the end.

Against this background, our model is premised on the assumption that, before negotiation can start, agents need to be jointly committed towards a state. According to the definition given in Section 2, a joint commitment that a group gr_i holds towards ϕ evokes a shared mental state in which, among other conditions, each member believes that ϕ is possible. In turn, this rests on the agents' belief that gr_i: (i) either can achieve ϕ directly; or (ii) can acquire the ability to achieve ϕ [23]. To say that a joint commitment must reflect either condition (i) or (ii) conveys the idea that, as long as a group has committed itself to achieve a state, it has committed itself to find the means to bring about that state, and each member's believing that there are no such means, either inside or outside the group, would contradict their joint commitment. On the one hand, should condition (i) be satisfied, then the agents would be aware of at least one plan that gr_i can perform to attain ϕ. On the other, should condition (ii) be satisfied, the agents would be aware of at least one plan that gr_i can perform to get closer to ϕ. This means that the agents believe they can discover how to achieve ϕ and eventually come up with a plan for ϕ. However, once a joint commitment has been formed, the agents may update their mental states and come to believe they are unable not only to fulfil their commitment directly but also to discover how to fulfil it. For example, they may overestimate their cognitive abilities and, ultimately, find out that the fulfilment of their joint commitment is beyond their abilities. In this case, no negotiation occurs since no plan is proposed as a potential candidate for being agreed upon within the group. Therefore, the minimum condition required in order for negotiation to take place is that at least one agent is aware of at least one plan that the group can perform to fulfil its joint commitment. This minimum condition is captured by Assumption 1.

Assumption 1. Given a group of agents jointly committed to achieving ϕ, a state will follow in which the group will hold its commitment iff at least one of the agents maintains a belief about a plan that the group might perform to attain ϕ:

$$\models \forall gr_i,\ \forall t_i, t_j\ (t_j>t_i)\ \textit{J-COMM}\ (gr_i, \phi(t_j))\ (t_i) \supset \exists t_k\ (t_i < t_k < t_j) \text{ s.t. } [\textit{J-COMM}\ (gr_i, \phi(t_j))\ (t_i, t_k) \Leftrightarrow \exists a_i \in gr_i, \exists e_i \text{ s.t. } \textit{Bel}(a_i, \textit{plan}(gr_i, e_i, \phi(t_j)))(t_k)]$$

Informally, Assumption 1 means that the agents will not keep their joint commitment forever. They will eventually drop their commitment unless, before the time the commitment is to be satisfied, at least one member of the group comes to

fully represent in its mind a potential candidate solution (i.e., a plan) for fulfilling it[4]. This allows negotiation to start. In fact, the agent's mental representation of a plan is the first step towards the generation and communication of a proposal to the other members of the group. However, before a proposal is forwarded, a *practical judgement* needs to be generated. To this end, the agent will keep trying to discover other possible alternative plans the group may perform [30]. Ultimately, the agent will typically come up with the belief that either there is only one possible plan that the group can execute, or that there are other alternative plans. In the former case, the agent will generate a *necessity-based* practical judgement: namely, the belief that, unless a specific plan is performed, the group cannot fulfil its joint commitment. In the latter case, the agent will have to evaluate and make a choice among the possible plans [2, 10, 16, 30]. Ultimately, this choice leads the agent to form a *preference-based* practical judgement: namely, the agent's preference that the group performs a plan, among a set of alternative feasible ones, based on the belief that the preferred plan will most satisfactorily enable the group to realise its joint commitment [23]. The generation of necessity- and preference-based judgements is formalised in Assumption 2.

Assumption 2. Should an agent hold a belief about a feasible plan, it will also generate a practical judgement:

$\models \forall gr_i, \forall t_i, t_j\ (t_j > t_i)$ *[J-COMM* $(gr_i, \phi(t_j))\ (t_i) \wedge \exists a_i \in gr_i, \exists e_i$ s.t. $Bel(a_i, plan(gr_i, e_i, \phi(t_j)))(t_i)] \supset Bel(a_i, (\phi(t_j) \Leftrightarrow <plan(gr_i, e_i, \phi)>(t_j)))(t_i) \vee \forall e_j\ (e_j \neq e_i)$ *[*$Bel(a_i, plan(gr_i, e_j, \phi(t_j))) \supset Pref(a_i, <plan(gr_i, e_i, \phi)>(t_j), <plan(gr_i, e_j, \phi)>(t_j))]\ (t_i) \vee \exists e_k\ (e_k \neq e_i)$ s.t. *[*$Bel(a_i, plan(gr_i, e_k, \phi(t_j)))(t_i) \wedge \forall e_j\ (e_j \neq e_k)$ *[*$Bel(a_i, plan(gr_i, e_j, \phi(t_j))) \supset Pref(a_i, <plan(gr_i, e_k, \phi)>(t_j), <plan(gr_i, e_j, \phi)>(t_j))]\ (t_i)$*]*

Informally, Assumption 2 means that if an agent is aware of a feasible plan for fulfilling the group's joint commitment, that agent will: (i) either believe that that plan is the only feasible one; or (ii) express the preference that the group performs that or another feasible plan among alternative feasible ones. Assumption 2 expresses the first two basic components of what is known as the agent's *social practical inference*. By this we mean the structure of the reasoning process that a member of a jointly committed group undertakes in order to give an answer to a practical problem [2, 10, 33]. On the one hand, the first conjunct of the antecedent of the above material implication represents the *major motivational premise* of the agent's inference, namely the joint commitment (and the agent's intention [23]) towards a state. On the other, the consequent suggests possible instantiations of the *minor doxastic premise* of the agent's inference. Typically, the minor premise contains a practical judgement suggesting a means for satisfying the joint commitment expressed within the major premise. The role of this judgement is to trigger the cognitive path leading the agent from the prior intention to achieve a state to a conclusive intention favouring the performance of a plan that brings about that state. This intention represents the *conclusion* of the agent's inference. While Assumption 2 was intended to formalise

[4] A solution to a practical problem may also be *emergent* from a path of *partial* plans. However, the emergent properties of negotiation fall outside the scope of this work.

the cognitive link between the two first inferential premises, Assumptions 3 and 4 formalise the generation of the conclusive inferential intention.

Assumption 3. Should an agent hold a necessity-based practical judgement favouring a plan, it will also generate the intention that the group performs that plan:

$\models \forall gr_i, \forall t_i, t_j\ (t_j > t_i)$ *[J-COMM* $(gr_i, \phi(t_j))\ (t_i) \wedge \exists a_i \in gr_i, \exists e_i$ s.t. $(Bel(a_i, (\phi(t_j) \Leftrightarrow <plan(gr_i, e_i, \phi)>(t_j)))(t_i)] \supset Int(a_i, <plan(gr_i, e_i, \phi)>(t_j))(t_i)$

Informally, Assumption 3 means that whenever an agent believes that there is only one plan that the group can perform to fulfil its commitment, it will conclude its social practical inference by generating the intention favouring that plan. However, the agent may also believe there is more than one feasible plan and, among the feasible ones, it may express a preference. Should an agreement have already been reached, the agent may decide to compromise over its preferences for the sake of the group (see section 4.1). However, when an agreement is still to be made, we expect the agent who expresses a preference for a plan to stick to it and generate an intention favouring that plan. The transformation of a preference-based practical judgement into the corresponding inferential intention is formalised in Assumption 4.

Assumption 4. Before an agreement is made, should an agent hold a preference-based practical judgement favouring a feasible plan, it will also generate the intention that the group performs that plan:

$\models \forall gr_i, \forall t_i, t_j\ (t_j > t_i)$ *[J-COMM* $(gr_i, \phi(t_j))\ (t_i) \wedge \neg\exists e_i$ s.t. *J-COMM*$(gr_i, <plan(gr_i, e_i, \phi)>(t_j))(t_i) \wedge \exists a_i \in gr_i, \exists e_i$ s.t. $[Bel(a_i, plan(gr_i, e_i, \phi(t_j)))(t_i) \wedge \forall e_j\ (e_j \neq e_i)\ [Bel(a_i, plan(gr_i, e_j, \phi(t_j))) \supset Pref(a_i, <plan(gr_i, e_i, \phi)>(t_j), <plan(gr_i, e_j, \phi)>(t_j))]\ (t_i)] \supset Int(a_i, <plan(gr_i, e_i, \phi)>(t_j))\ (t_i)$

Informally, Assumption 4 means that whenever an agent prefers a plan over a range of alternative feasible ones, and an agreement is still to be reached within the group, then the agent will form the intention favouring the preferred plan. Both this preference-based intention and the necessity-based one pave the way for subsequent socio-cognitive processes [34]. In fact, in order for a group to successfully undertake negotiation, the members' inferential intentions must be socially interconnected in such a way that an agreement can be reached. This motivates a subsequent inter-agent *coordination* process. Once one of the agents has generated an intention favouring some plan, we expect that agent to generate the intention to bring about a state in which all its acquaintances know about that plan [36]. This is expressed in Assumption 5.

Assumption 5. Should an agent intend that the group performs a plan, it will also intend to bring about a state where every member is aware of this:

$\models \forall gr_i, \forall t_i, t_j\ (t_j > t_i)$ *[J-COMM*$(gr_i, \phi(t_j))(t_i) \wedge \exists a_i \in gr_i, \exists e_i$ s.t. $Int(a_i, <plan(gr_i, e_i, \phi)>(t_j))(t_i)] \supset \forall a_j \in gr_i\ Int(a_i, Bel(a_j, Int(a_i, <plan(gr_i, e_i, \phi)>(t_j))))\ (t_i)$

Informally, an agent's intention to make the group perform a plan leads the agent to exert social influence upon its acquaintances' mental states. In Assumption 5 this has been formalised through a nested modal operator: *the intention about somebody's belief about somebody's intention*. In most cases, the agent attempts to impact upon the other members' mental states by sending a message and letting them know about the plan it favours. This plan will be a candidate for being moved up to an agreed-upon plan status.

The intention to let somebody know something can be regarded as an instantiation of a more general attitude: the intention to make somebody adopt a mental attitude [24]. This is a key construct that lies at the heart of most social processes and inter-agent social behaviours. In fact, it can be seen as the cognitive source of a variety of social influence processes that agents exert in order to impact upon each other's mental states. If social influence is successful, the agent who is subjected to it will typically change its mental state and adopt new *socially motivated* mental attitudes. These are attitudes that are inherently motivated by social behaviour and rooted in the agents' disposition to represent each other in intentional terms [14]. In particular, when social influence rests on an agent's intending to let another know something, the typical outcome is the latter's adoption of a socially motivated belief.

In their attempts to exert social influence, agents may fail to let others know about the plans they favour. Agents may simply lack an appropriate communication channel, or speak different languages. In this case, agents fail to influence one another, and the agreement is procrastinated, if not hampered. However, communication may also be effective and social influence may succeed. In this case, once the agents have come to know about one of their acquaintances' proposals, they will update their beliefs and evaluate the message received. Each agent will then act in differing ways depending on the extent to which the proposal is consistent with its own beliefs, intentions, preferences, etc. Specifically, should an agent keep its commitment to the group, it may react in two different ways. First, it may agree with its acquaintance and endorse the intention that the group performs the proposed plan. Second, it may disagree, and reject the proposal [20, 25]. In this case, should the agent support a different plan, it will also generate the corresponding intention to make all the other members aware of this (Assumption 5). To this end, the agent will typically generate and communicate a *counter-proposal* [16, 28]. The process then iterates with all the other agents' evaluating the counter-proposal and, ultimately, with their rejection or acceptance.

The last assumption we want to make about agents' behaviour is concerned with the generation of a final agreement. In this respect, we note that, even after a number of proposals and counter-proposals have been forwarded, negotiation may still fail. The agents may simply be unable to reach an agreement, due to some irreconcilable differences. However, negotiation may also succeed and end up with an agreement based on a joint commitment to jointly performing a plan. In this case, before a conclusive joint commitment is generated, we expect all the agents to endorse the same intention about the plan to be performed[5]. However, all the agents' sharing the same intention is a necessary but not sufficient condition for an agreement to be

[5] Note that we generalise over specific mechanisms of agreement-generation (e.g. democratic voting mechanisms; majority rules; appealing to authority) which do not necessarily reflect each member's sharing the same intention [32].

reached. In fact, for example, the agents may be unable to establish a mutual belief that they share the same intention, and this does not enable them to generate a joint intention, and therefore a joint commitment supporting a common plan (section 2) [23, 36]. Therefore, if we want to be able to say that negotiation concludes successfully when the agents favour the same plan, we need to make another assumption about agents' behaviour. More precisely, we need to assume that when agents share the same intention supporting a plan, they are also successful in establishing a shared mental state from which a joint commitment to performing that plan ensues. This is formalised in Assumption 6.

Assumption 6. If all the group members share the intention that the group performs a given plan, they will become jointly committed to performing that plan:

$\models \forall gr_i, \forall e_i, \forall t_i, t_j\ (t_j > t_i)\ [\text{J-COMM}(gr_i, \phi(t_j))(t_i) \wedge \forall a_i \in gr_i\ \text{Int}(a_i, \langle \text{plan}(gr_i, e_i, \phi) \rangle (t_j))(t_i)] \supset \text{J-COMM}(gr_i, \langle \text{plan}(gr_i, e_i, \phi) \rangle (t_j))(t_i)$

Assumption 6 formalises the core idea underpinning our model, namely the fact that negotiation can be regarded as a transformation of commitments. When successful, negotiation generates an agreement on a joint plan [28]. An agreement is a composite concept that reflects what practical judgements the agents have brought about via practical reasoning and to what extent the agents have compromised with one another over their own views and preferences. Furthermore, an agreement reflects a joint commitment to acting in accordance with the agreed-upon plan. Coming to an agreement thus transforms a joint commitment towards a state into a joint commitment to performing a plan for achieving that state. This transformation of commitments is what constitutes the essence of negotiation.

4 Properties of the Model

Drawing on the classical axiomatic-deductive methodology for theory building [27], in this section properties of the model will be discussed and formalised. In doing so, by exploring some major problems occurring in real-world negotiations, it will be shown how it is possible to make a step towards a cross-fertilisation among different disciplines, and in particular mainstream DAI and organization theory.

4.1 Cognition at the Individual and the Joint Level

The objective of this section is to examine properties concerning the relationship between the negotiated agreement and the individual mental attitudes. First, we note that at the heart of an agreement lie the individual agents' intentions. Should negotiation be successful, not only will the agents share an intention towards an end, but they will also share an intention towards the means to secure that end.

Property 1. Agreement rests on and transcends individual intentions:

$\models \forall gr_i, \forall e_i, \forall\ t_i,\ t_j\ (t_j>t_i)\ [J\text{-}COMM(gr_i,\ \phi(t_j))(t_i) \wedge J\text{-}COMM(gr_i, <plan(gr_i,\ e_i,\ \phi)>(t_j))(t_i)] \supset \forall a_i \in gr_i\ [Int(a_i,\ \phi(t_j))\ (t_i) \wedge Int(a_i, <plan(gr_i,\ e_i,\ \phi)>(t_j))\ (t_i)]$

This property follows from the definition of joint commitment (see section 2 and [23]). Informally, if any two agents come to an agreement, they both endorse the same intention towards a state and a plan for achieving that state. Conversely, sharing identical intentions does not imply that negotiation has been carried out and an agreement has been reached (i.e., the implication above is unidirectional). The reason for this is quite simple. Having identical intentions is not sufficient for a joint commitment to take place [23]. As a result, the conjunction of the agents' intentions towards a state and a plan does not imply the joint commitments towards that state and that plan, which in our formalisation represent, respectively, the pre-condition and the ultimate outcome of negotiation.

While Property 1 is concerned with the implications of an agreement in terms of the agents' intentions, Property 2 explores the implications of sharing identical intentions in terms of the agents' practical judgements. Even though, in order to reach an agreement, agents are expected to change their mental states until they share the same intention about a plan, nonetheless they are not required to adapt their preferences to each other in a consistent manner. In fact, they might agree and still have divergent preferences and personal views as to what is the most appropriate plan the group should perform.

Property 2. Agreement about a plan does not rest on identical preference-based practical judgements favouring that plan:

$\not\models \forall gr_i, \forall e_i,\ \forall\ t_i,\ t_j\ (t_j>t_i)[\ J\text{-}COMM(gr_i,\ \phi(t_j))(t_i) \wedge J\text{-}COMM(gr_i,\ <plan(gr_i,\ e_i,\ \phi)>(t_j))(t_i)] \supset \forall a_i \in gr_i\ \forall e_j\ (e_j \neq e_i)\ [Bel(a_i,\ plan(gr_i,\ e_j,\ \phi(t_j))) \supset Pref(a_i,\ <plan(gr_i,\ e_i,\ \phi)>(t_j),\ <plan(gr_i,\ e_j,\ \phi)>(t_j))]\ (t_i)$

The antecedent does not imply the consequent because, should the antecedent be satisfied, and the agents share the same intentions towards a state and the performance of a plan (Property 1), one of the agents might still prefer that the group performs another plan. This follows from the axiomatisation of intentions and the conditions imposed on the intention-accessibility relation (section 2). More specifically, intentions are not taken to be constrained by any intention-preference consistency schema.

The fact that the agents who agree about a plan do not necessarily hold a preference-based practical judgement favouring that plan, has an interesting implication concerning one of the key problems found in real-world negotiations: *compromising* and *intention reconsideration*. Even though the agents keep their views and personal preferences, they nonetheless may need to compromise and drop their individual intentions for the sake of the group [32, 35]. Let us suppose that, at time t_i, group gr_i has successfully carried out a negotiation regarding a state ϕ to be achieved at t_j $(t_j>t_i)$. Formally, we have: $J\text{-}COMM(gr_i,\ \phi(t_j))(t_i) \wedge J\text{-}COMM(gr_i,\ <plan(gr_i,\ e_i,\ \phi)>(t_j))(t_i)$. Furthermore, suppose that at $t_h<t_i$ one of the agents, say a_i, generated a preference-based practical judgement favouring a plan, say e_j, that is different from

the agreed-upon plan e_i. Formally, we have: $\exists\, e_j\ (e_j \neq e_i)$ s.t. $[Bel(a_i, plan(gr_i, e_j, \phi(t_j)))(t_h) \wedge \forall\, e_h\ (e_h \neq e_j)\ [Bel(a_i, plan(gr_i, e_h, \phi(t_j))) \supset Pref(a_i, <plan(gr_i, e_j, \phi)>(t_j), <plan(gr_i, e_h, \phi)>(t_j))]\ (t_h)]$. Now, according to Property 1, agreement on e_i implies $Int(a_i, <plan(gr_i, e_i, \phi)>(t_j))(t_i)$. However, according to Assumption 4, if an agent holds a preference-based practical judgement before an agreement is made, the agent will also generate the corresponding intention favouring that judgement, that is, $Int(a_i, <plan(gr_i, e_j, \phi)>(t_j))(t_h)$. Thus, if the agent is to comply with the agreement, an intention reconsideration must occur leading the agent to drop the intention based on its own preference-based practical judgement in order to endorse the intention favoured by the group. Endorsing a socially motivated intention to the detriment of an internally motivated one thus represents the cognitive implication of the compromises that agents are often required to make with one another over their own preferences to get to a final agreement and stick to it.

Compromising and intention reconsideration have often been regarded as inherently intertwined with the role and implications of *conflict* in negotiation [28]. In this respect, it has been argued that negotiation is a response to conflict [17]. However, it is worth noting that, even though particularly common in most real-world scenarios, conflict cannot be seen as a necessary condition that motivates agents into negotiation. In the same vein, compromising and intention reconsideration do not underpin every real-world negotiation. For example, the agents may share the same preferences and intentions without being aware of this. In this case, they need to negotiate in order to find out they all agree on what the group should do. Furthermore, since here no conflict exists between the agents, they are not expected to compromise with one another over their own preferences, nor will they re-consider their individually motivated intentions. On a more general level, rather than simply as a response to conflict, negotiation occurs whenever the agents do not know whether or not they share the same preference/intention as to how to fulfil their joint commitment. Negotiation thus is intended to make the agents realise either that they already converge or that they need to compromise with one another to overcome their divergence.

4.2 Private Information, Bounded Rationality and Informational Asymmetries

A key problem in most real-world negotiations is the uncertainty and ambiguity of the information needed to reach an agreement. There are two main reasons for this. First, in most circumstances different agents have differing relevant *private information* before an agreement is reached. As a result of this, the information that is needed to reach an agreement tends to be localised and dispersed among the agents. Second, in most real-world scenarios agents are *boundedly rational* [30]. They have limited cognitive ability, imperfect communication skills and their natural languages are imprecise. In particular, agents cannot solve arbitrarily complex problems exactly, costlessly and instantaneously, they cannot process all the information they have simultaneously and accurately, they cannot communicate with one another freely and perfectly, and the understanding of messages is often flawed. More specifically, determining what information a message conveys, what message should be forwarded to whom, and with what methods messages should be forwarded becomes an

overwhelmingly large and complex problem. Because information is localised and dispersed, and because no one has the cognitive ability to make all the calculations needed to retrieve information, the agents cannot account for all the relevant information needed to determine the best use of resources and the appropriate adaptations. Thus, it is upon the link between informational dispersion and the agents' bounded rationality that rests the problem of *informational inaccuracy and asymmetry*, which, in turn, represents a major obstacle that interferes with the possibility of reaching a mutually beneficial agreement [30].

The following property is precisely intended to give a formalisation of the fact that, in most real-word negotiations, the agents' doxastic representations of their social environment are inherently ambiguous. More specifically, Property 3 conveys the idea that the negotiating agents' beliefs about each other's mental attitudes are not deterministically accurate. They are not inevitably true in the same way as they are not inevitably false.

Property 3. Agents are boundedly rational in generating and updating their beliefs about each other's intentions:

$$\nvdash \forall\, gr_i,\ \forall\, a_i,\ a_j \in gr_i,\ \forall\, e_i,\ \forall\, t_i,\ t_j\ (t_j>t_i)[\text{J-COMM}(gr_i,\ \phi(t_j))(t_i) \wedge \text{Bel}(a_j,\ \text{Int}(a_i,<\text{plan}(gr_i,\ e_i,\phi)>(t_j)))(t_i)] \supset \text{Int}(a_i,<\text{plan}(gr_i,\ e_i,\phi)>(t_j))(t_i)$$

The property follows from the fact that beliefs are not taken to be constrained by the "knowledge axiom" and, correspondingly, the belief-accessibility relation is not taken to be reflexive (Section 2). Therefore, what an agent believes might turn out to be false. More specifically, an agent may inaccurately believe that another holds an intention that it actually does not.

Property 3 enables to highlight two complementary conceptualisations of the agent's bounded rationality. First, should agent a_i generate its belief at time t_i, Property 3 means that this may be inaccurately formed. In fact, the agent may mistakenly believe that another holds an intention. This is consistent with what has been argued by mainstream organisation and cognitive science, namely that the (human) agent typically forms mental models of its environment by using imperfect representations that simplify the complexity of spatial, temporal and causal relationships [30]. Second, should agent a_i generate its belief at some time $t_j < t_i$, Property 3 means that, not only may beliefs be inaccurately formed, but the agent may also be unable to update them once formed. In fact, in this case, the agent may keep maintaining a belief that another holds an intention, even though the latter does not hold that intention. Here, the agent's limited ability to update its beliefs does not allow it to discover every change in its environment that might falsify its cognitive representations formed at an earlier stage. Thus, even beliefs that were accurately formed may subsequently turn out to be false during the course of negotiation as a result of changes in the environment and the agent's limited ability to account for them.

4.3 Conflicting Interests and Opportunistic Behaviour

The property discussed in this section is concerned with another central problem of real-world negotiations: *opportunistic behaviour* and the related issue of *motivation.* Not only have real agents limited cognitive ability. They also have their own private interests, which are rarely perfectly aligned with the interests of the other agents with whom they need to interact [13]. Divergence of interests, together with bounded rationality and information specificity, introduces the possibility of opportunistic behaviour. Because agents are boundedly rational, they suffer from informational distortions that might prevent them from reaching a mutually beneficial agreement. However, bounded rationality and informational ambiguity can also be *exploited* by the agents to opportunistically misrepresent or even refuse to reveal relevant private information. Typically, agents might exploit their counter-parts' bounded rationality in order to obtain a unilateral advantage and seize a greater share of the fruits of negotiation for themselves [21]. Correspondingly, the motivation problem is to ensure that the various agents involved in negotiation willingly do their parts in the whole undertaking, both communicating information accurately to allow the right agreement to be reached and acting as they are expected to act within the group.

Opportunistic behaviour and motivation are inherently related to each other. In fact, should the agents not be sufficiently motivated to act in a way that is beneficial to the whole group, they might behave opportunistically and hide relevant private information, or even alter it in an effort to have their own interests and objectives satisfied at the expense of the others'. This source of inefficiency is often called *adverse selection*, conveying the idea that one party behaves in a way that is detrimental, adverse to the interests of the other party [1]. Since agents are aware of the fact that their acquaintances cannot possibly have all the relevant information that is needed to accurately evaluate the counter-part's behaviour, they may be induced to misbehave and misrepresent their private information in order to have their interests fulfilled more quickly or efficiently. In particular, an agent may intend that another mistakenly perceives that it has a specific intention about a potential agreement, perhaps because this may induce the latter to act in a manner that is beneficial to the former [21]. Or, an agent may intend to hide its own strategy to another agent, thus deliberately forwarding wrong messages to the latter in an attempt to make it generate inaccurate beliefs. These forms of opportunistic misbehaviour are captured in the following property.

Property 4. Agents may opportunistically mislead each other into thinking that they maintain intentions they actually do not:

$$\nvDash \forall gr_i,\ \forall a_i,\ a_j \in gr_i,\ \forall e_i,\ \forall t_i,\ t_j\ (t_j>t_i)[\text{J-COMM}(gr_i,\ \phi(t_j))(t_i) \wedge \text{Int}(a_i,\ \text{Bel}(a_j,\ \text{Int}(a_i,\ <\text{plan}(gr_i,\ e_i,\phi)>(t_j))))(t_i)] \supset \text{Int}(a_i,<\text{plan}(gr_i,\ e_i,\ \phi)>(t_j))(t_i)$$

The antecedent does not imply the consequent because, although the antecedent is satisfied, agent a_i might intend that the group performs a plan, say e_j, different from e_i. In this case, the consequent would not be satisfied. More generally, in our axiomatisation of agents' mental attitudes (section 2), the nested operator $\text{Int}(a_i, \text{Bel}(a_j, \text{Int}(a_k,\ \phi)))(t_i)$, expressing the intention about somebody's belief about somebody's

intention, for arbitrary a_i, a_j, and a_k, is not taken to imply $Int(a_k, \phi)(t_v)$. Thus, an agent's intention that another believes it has an intention favouring a plan does not entail that it really intends to favour that plan.

Intuitively, what Property 4 suggests is that the inaccuracy of agents' beliefs may be the result of somebody else's opportunistic behaviour. As we noted above, this self-interested misbehaviour leading to adverse selection is a form of *pre-contractual* opportunism that arises because of the private information that boundedly rational agents have *before* they reach an agreement [1]. However, besides this, there is another form of self-interested misbehaviour, known as *moral hazard*, that reflects the agents' proclivity to behave opportunistically *after* an agreement has been made [26]. The cognitive roots of this form of post-contractual opportunism can be found in the dynamic implications of the agents' bounded rationality. Not only have agents inaccurate information about the world; they are also not perfectly far-sighted [30]. In particular, their ability to make agreements is limited by the existence of unforeseen circumstances and by the costs and difficulty of deciding in advance what would be appropriate to do in every foreseeable contingency. Even in the extreme case, where there is no private information before an agreement is made, there may be inadequate information afterwards to tell whether the terms of the agreement have been honoured, or acquiring that information may be costly. As a result, agreements are inevitably incomplete and imperfectly specified, so that the agents involved can exploit loopholes to gain an advantage over one another [26]. In addition, as shown with the problem of adverse selection, actions that have efficiency consequences are not freely observable and so the agent taking them may choose to pursue its private interest at others' expense [21].

Now, having modelled negotiation and the generation of an agreement in terms of a transformation of joint commitments, we can easily show how this form of moral hazard can be accounted for in our formalisation. In fact, our conception of negotiation can be seen as implicitly reflecting an underpinning form of meta-agreement, namely an *agreement about reaching an agreement*. In this respect, being jointly committed to negotiating as to how to attain some state means to agree to attain that state collaboratively, and therefore to agree to eventually make an agreement about the appropriate plan. In the light of this, Property 4 can now be read from a different perspective that enables the problem of moral hazard to be accounted for. Because boundedly rational agents cannot foresee all the relevant circumstances, the agreements they can make are inevitably incomplete. This is true also of a meta-agreement concerning how to reach an agreement. That is, the prior joint commitment to engaging in negotiation cannot conceivably specify all the relevant circumstances that might arise during negotiation. This opens the incentive for agents with divergent interests to misbehave in such a way that their own private interests can be fulfilled with no mutual advantage for the others. Again, this form of misbehaviour is captured in our formalism by an agent's intending that another mistakenly perceives something, after a joint commitment has been formed. Since the way in which negotiation is to be undertaken cannot be completely specified in advance, the agents, once jointly committed, may be induced to misbehave and try to fulfil the joint commitment in a self-interested manner.

5 Conclusions

In this paper we used a BDI logic to formalise agents' behaviour and cognition during negotiation in multi-agent environments. In doing so, we have been motivated by two objectives. First, to place the study of negotiation on a more secure and formal footing. Second, to develop an empirically satisfactory theory, comprehensive enough to account for a number of problems occurring in real-world negotiations. To this end, by formalising a number of assumptions, our approach was first, to synthetically generalise over specific strategies and tactics that agents might use in different domains; second, to analytically specify the key cognitive and social processes that underpin most forms of real-world negotiations. In this respect, our perspective differs from computational (e.g. [16]) as well as economic and game-theoretic (e.g. [28]) approaches to negotiation, whereby a range of tactics, inter-agent behavioural patterns, decision and preference functions are formally specified and empirically evaluated. Conversely, our work is most closely related to the formalisms described in [20], [25] and [36], where negotiation is modelled in terms of the agents' decision-making apparatus. However, with respect to these works, our focus has been more on the representation of practical inferential processes in terms of a transition between mental attitudes, as well as on the analysis of the cognitive foundations of inter-agent social influence processes [6, 34].

In developing our formalisation, we brought some of the major research questions in social sciences to bear on the methods and analytical tools advocated by mainstream computer science and DAI. For example, we attempted to formalise such problems as the agent's bounded rationality, adverse selection and moral hazard, using a computational BDI logic. Furthermore, we worked towards a cross-fertilisation among research questions suggested by different theoretical perspectives. For example, we studied the problem of informational asymmetry, typically addressed by economists and organisational scientists [1, 21, 26], by focusing on the agent's cognitive architecture, which is a conceptual category imported from DAI [11, 35]. In this vein, by building a new conception of negotiation upon the interconnections among different disciplines, in this paper we took some steps towards a unified theoretical and methodological paradigm for modelling the cognitive foundations of pluralism in MASs.

References

1. G. Akerlof. The market for lemons: Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84: 488-500, 1970.
2. R. Audi. A theory of practical reasoning. *American Philosophical Quarterly*, 19(1): 25-39, 1982.
3. J. Bell. Changing attitudes. In M. Wooldridge and N. R. Jennings, editors. *Intelligent Agents*, Springer, Berlin, pp. 40-55, 1995.
4. J. Bell and Z. Huang. Dynamic goal hierarchies. In L. Cavedon, A. Rao and W. Wobcke, editors, *Intelligent Agent Systems: Theoretical and Practical Issues*, Springer-Verlag, Berlin, pp. 88-103, 1997.
5. R.M. Burton and B. Obel. The validity of computational models in organisation science: From model realism to purpose of the model. *Computational and Mathematical Organisation Theory*, 1(1): 57-71, 1995.
6. K. M. Carley. The value of cognitive foundations for dynamic social theory. *Journal of Mathematical Sociology*, 14(2-3):171-208, 1989.
7. K. Carley and M. J. Prietula, editors. *Computational Organisation Theory*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.

8. C. Castelfranchi. Commitments: From individual intentions to groups and organisations. In *Proceedings of the First International Conference on Multi-Agent Systems*, AAAI Press and MIT Press, San Francisco, CA, pp. 41-48, 1995.
9. B. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, MA, 1980.
10. S. Clarke. *Practical Inferences*. Routledge and Kegan Paul, London, 1985.
11. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(3): 213-261, 1990.
12. P. R. Cohen and H. J. Levesque. Teamwork. *Nous*, 25(4): 487-512, 1991.
13. M. Crozier and E. Friedberg. *L'Acteur et le Système*. Éditions du Seuil, Paris, 1977.
14. C. Dennett. *The Intentional Stance*. The MIT Press, Cambridge, MA, 1987.
15. B. Dunin-Keplicz and R. Verbrugge. Collective commitments. In *Proceedings of the Second International Conference on Multi-Agent Systems*, Menlo Park, CA, pp. 56-63, 1996.
16. P. Faratin, C. Sierra and N. R. Jennings. Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems*, 24(3-4): 159-182, 1998.
17. L. Greenhalgh and D. I. Chapman, Joint decision-making. The inseparability of relationships and negotiation. In R. M. Kramer and D. M. Messick, editors. *Negotiation as a Social Process*. SAGE Publications, Thousand Oaks, CA, pp. 166-185, 1995.
18. J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, NY, 1962.
19. D. Kinny, M. Ljungberg, A. S. Rao, E. Sonenberg, G. Tidhar and E. Werner. Planned team activity. In C. Castelfranchi and E. Werner, editors, *Artificial Social Systems – Selected Papers from the Fourth European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW-92*, vol. 830, Springer-Verlag, Heidelberg, pp. 226-256, 1992.
20. S. Kraus, K. Sycara and A. Evenchil. Reaching agreements through argumentation: A logical model and implementation. *Artificial Intelligence* 104: 1-69, 1998.
21. P. Milgrom and J. Roberts. Bargaining costs, influence costs, and the organization of economic activity. In J. Alt and K. Schepsle, editors. *Perspectives on Positive Political Economy*, Cambridge University Press, Cambridge, MA, 1990.
22. P. Panzarasa and N. R. Jennings. The organisation of sociality: A manifesto for a new science of multi-agent systems. In *Proceedings of the Tenth European Workshop on Multi-Agent Systems (MAAMAW-01)*, Annecy, France, 2001
23. P. Panzarasa, N. R. Jennings, and T. J. Norman. Formalising collaborative decision-making and practical reasoning in multi-agent systems. *Journal of Logic and Computation*, 11(6): 1-63, 2001.
24. P. Panzarasa, N. R. Jennings, and T. J. Norman. Social mental shaping: Modelling the impact of sociality on the mental states of autonomous agents. *Computational Intelligence*, forthcoming, 2001.
25. S. Parsons, C. Sierra and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3): 261-292, 1998.
26. M. Pauly. The economics of moral hazard. *American Economic Review*, 58: 31-58, 1968.
27. K. R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1959.
28. H. Raiffa. *The Art and Science of Negotiation*. Harvard University Press, Cambridge, MA, 1982.
29. A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3): 293-342, 1998.
30. H.A. Simon. *Administrative Behavior*. 3rd Edition, Free Press, New York, NY, 1976.
31. M. P. Singh. *Multiagent Systems: A Theoretical Framework for Intentions, Know-how, and Communications*. Springer-Verlag, Berlin, 1995.
32. A. Strauss. *Negotiations: Varieties, Contexts, Processes, and Social Order*. Jossey-Bass, San Francisco, CA, 1978.
33. R. Tuomela. *A Theory of Social Action*. Reidel Pub., Boston, 1984.
34. K. E. Weick. Cognitive processes in organizations. *Research in Organization Behavior*, JAI Press, 1: 41-74, 1979.
35. M. Wooldridge. *Reasoning About Rational Agents*. The MIT Press, Cambridge, MA, 2000.
36. M. Wooldridge and N. R. Jennings. Cooperative problem solving. *Journal of Logic and Computation*, 9(4): 563-592, 1999.

H. von Wright. *The Logic of Preference*. Edinburgh University Press, Edinburgh, 1963.