Multi-Agent Negotiation: Logical Foundations and Computational Complexity

P. Panzarasa
University of London
p.panzarasa@qmul.ac.uk

K. M. Carley
Carnegie Mellon University
Kathleen.Carley@cmu.edu

Abstract

This paper is concerned with the use of logic-based formalisms for multi-agent negotiation. We begin by introducing a quantified multi-modal language that draws on and extends standard Belief-Desire-Intention (BDI) logics. Using this language, a number of properties concerning the behavior and cognition of negotiating agents will be examined on a proof-theoretic basis. We then concentrate on the computational complexity of a fundamental problem that arises in multi-agent negotiations – the problem of determining whether negotiation guarantees coordination among interdependent agents. To this end, we introduce a series of progressively more sophisticated negotiation protocols, and consider how computational complexity varies, depending on the properties of these protocols.

Contact:
Pietro Panzarasa
Centre for Business Management
Queen Mary, University of London
Mile End Road
London, E1 4NS, U.K.

Tel: +44 (0)20-7882-6984
Fax: +44 (0)20-7882-3615
Email: p.panzarasa@qmul.ac.uk

Key Words: coordination, agent communication, protocol, computational complexity.

Multi-Agent Negotiation: Logical Foundations and Computational Complexity

Pietro Panzarasa and Kathleen M. Carley

In this paper, negotiation is regarded as a coordination mechanism for governing the diversity of interests and knowledge among interdependent autonomous agents. A quantified multi-modal logical language is developed for reasoning about and representing agents' mental attitudes and behaviors. Drawing on this language, negotiation is formalized 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. We then define protocols for negotiation, and consider the problem of whether a particular protocol guarantees that agreement between negotiating agents will be reached. We investigate how the computational complexity of this problem varies, depending on the properties of the protocol.

Preliminaries

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. 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. Time is taken to be composed of points and, for simplicity, is assumed to be discrete and linear [Emerson, 1990]. Terms come in six sorts denoted by mutually disjoint sets of symbols: agents (denoted by a_i, a_j, …), groups of agents (defined as non-empty subsets of the set of agents and denoted by gr_i, gr_j, …), time points (denoted by t_i, t_j, …), temporal intervals (defined as pairs of time points and denoted by i_i, i_j, …), actions (denoted by e_i, e_j, …) and generic objects in the environment (denoted by o_i, o_j, …). Membership in groups is expressed through the "$\in$" operator: $a_i \in gr_i$ means that the agent denoted by a_i is a member of the group denoted by gr_i. In addition, the language includes first-order equality, the classical connectives "$\neg$" (not) and "$\vee$" (or) and the universal quantifier "$\forall$". The remaining classical connectives and existential quantifier are assumed to be introduced as abbreviations, in the obvious way. Finally, the alphabet of the logic contains the punctuation symbols ")", "(", "[", "]", and comma ",".

Operators for Reasoning about Actions and Mental Attitudes

To express the occurrence of an action in the world, the language includes the operator $Occurs(e_i)(t_i)$, which means that action e_i happens at time t_i. Actions may be performed by an individual agent or by a group of agents. 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 capture the notion of a state-directed action, we introduce the derived operators $plan(gr_i, e_i, \phi(t_j))(t_i)$ and $plan(a_i, e_i, \phi(t_j))(t_i)$. At time t_i action e_i is a plan for group gr_i (or agent a_i) to make ϕ true at t_j $(t_j>t_i)$ iff: (a) e_i will occur sometime before t_j; (b) gr_i (or a_i) is the group (or agent) required to perform e_i; and (c) if e_i occurs, then ϕ will be satisfied afterwards at t_j [Panzarasa *et al*., 2001a]. The above definition of plans refers to actions that agents or groups *eventually* perform to satisfy certain states of the world. We also want to express the *past* execution of state-directed actions. To this end, we introduce the operators $<plan(gr_i, e_i, \phi)>(t_i)$ and $<plan(a_i, e_i, \phi)>(t_i)$. At time t_i, ϕ has been made true as a consequence of the performance of action e_i by group gr_i (or agent a_i) iff: (a) e_i occurred sometime in the past; (b) gr_i (or a_i) was the group (or agent) required to perform e_i; and (c) ϕ was satisfied afterwards at t_i as a consequence of the performance of e_i.

The logic is further enriched by a set of modal operators for reasoning about agents' mental attitudes [Carley and Newell, 1994]. Drawing on a fairly standard BDI framework, the operators $Bel(a_i, \phi)(t_i)$ and $Int(a_i, \phi)(t_i)$ 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 [Wooldridge, 2000]. Firstly, beliefs may concern facts of the world and can be nested. The formal semantics for beliefs are a natural extension of the traditional Hintikka's possible-worlds semantics [Hintikka, 1962]. 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 [Rao and Georgeff, 1998]. Secondly, intentions represent the states of the world that agents are "self-committed" to achieving or maintaining. 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 [Rao and Georgeff, 1998].

In addition to beliefs and intentions, agents can have preferences and commitments. Firstly, 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 [Panzarasa *et al.*, 2001]. Secondly, the operator $Comm(a_i, gr_i, e_i)(t_i)$ means that at time t_i agent a_i is committed towards group gr_i to performing action e_i. 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 $(t_j>t_i)$. At t_i agent a_i is socially committed towards group gr_i to making ϕ true at t_j $(t_j>t_i)$ 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.

Having defined individual agents' mental attitudes, we now turn to the cognitive properties of groups and introduce three modal operators for reasoning about joint mental attitudes. Firstly, $M\text{-}BEL\ (gr_i, \phi)(t_i)$ 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 [Fagin *et al.*, 1995]. Secondly, the operator $J\text{-}INT(gr_i, \phi)(t_i)$ captures the notion of *joint intention*. 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. Finally, the operator $J\text{-}COMM(gr_i, \phi(t_j))(t_i)$ allows us to represent a group's *joint commitment* to achieving a state of the world. 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.

Inter-Agent Communication and Agreement

Having introduced the general logical framework, we now formalize negotiation in terms of the type of dialogue and agreement among interdependent parties. In so doing, our analysis reflects two core ideas regarding the nature of negotiation. First, negotiation implies communication among identified parties: locutions are exchanged in an attempt to find ways of fulfilling the interests of the parties involved as much as possible. Second, negotiation is regarded as primarily aimed at generating an agreement among agents on which action to undertake. This agreement reflects the agents' joint commitment to acting according to a joint course of action. In this vein, an agreement can be regarded as a solution to the problem of how to effectively coordinate interdependent agents with conflicting interests, heterogeneous preferences and distributed knowledge. This view is consistent with the idea of negotiation as a searching process aimed at generating alternative courses of action that are potential solutions to a practical problem. The problem concerns what is to be done by a group of interdependent agents; the agreed solution reflects a conclusive joint commitment towards the means to secure the end.

A key role in the generation of an agreement is played by *social influence*. In fact, the coordination of interdependent agents implies processes of reciprocal modification of mental states and behaviors. For example, 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. An agent's social influence upon its acquaintances' mental states can be formalized through a nested modal operator: the intention about somebody's belief about somebody's intention. 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 [Panzarasa *et al.*, 2001b]. This is a key construct that lies at the heart of most social processes and inter-agent social behaviors. 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 motivated by social behavior and rooted in the agents' capabilities to represent each other in intentional terms.

Properties of Negotiation

Drawing on the classical axiomatic-deductive methodology for theory building, three sets of properties of negotiation can be derived and formalized. The first set is concerned with the relationship between the negotiated agreement and the individual agents' mental attitudes. Agreements rest on and transcend individual attitudes. If any two agents come to an agreement, they both endorse the same intention towards a plan for achieving a state. Nonetheless, sharing identical intentions does not imply that an agreement has been reached. Furthermore, agreements do not require the agents 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 the most appropriate plan that the group should perform. This has an interesting implication concerning one of the key problems of real-world negotiations: *compromising* and *intention reconsideration*. Since agents may not change their views and preferences, they may need to compromise and drop their individual intentions for the sake of the group [Wooldridge, 2000]. Endorsing a socially motivated intention to the detriment of an internally motivated one reflects the compromises that agents are often required to make with one another over their own preferences in order to get to an agreement and stick to it.

The second set of properties deals with another key problem of most real-world negotiations: the *asymmetry* and *inaccuracy* of the information needed to reach an agreement. On the one hand, 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 localized and dispersed among the agents. On the other, in most real-world scenarios agents are *boundedly rational* [Simon, 1976]. They have limited cognitive ability, imperfect communication skills and their natural languages are imprecise. As a result of the combined effects between informational asymmetry and inaccuracy, 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. This may represent a major obstacle that interferes with the possibility of reaching a mutually beneficial agreement.

The third set of properties is concerned with the problem of *opportunistic behavior*. Typically, agents have their own private interests, which are rarely perfectly aligned with the interests of the other agents with whom they need to interact. Divergence of interests, combined with bounded rationality and information specificity, introduces the possibility of opportunistic behavior [Williamson, 1964]. In fact, agents might exploit each other's limited cognitive abilities in order to obtain a unilateral advantage and seize a greater share of the fruits of negotiation for themselves. For example, they might opportunistically misrepresent or even refuse to reveal relevant private information. They might opportunistically mislead each other into thinking that they maintain intentions they actually do not. Or, once an agreement has been reached, they might try to fulfil their joint commitment in a self-interested manner.

The Computational Complexity of the Negotiation Problem

Drawing on the logical framework presented in the previous sections, we now explore the intrinsic computational difficulty of a fundamental problem that arises in multi-agent environments. This problem can be informally stated as follows: Given a set of interdependent agents, does negotiation guarantee that the agents' actions will be effectively and efficiently coordinated? To date, mainstream social sciences have addressed this issue primarily by focusing on the tactics and strategies for improving the outcomes of negotiation [Lewicki *et al.*, 2003]. However, the problem of how much computing power and/or resources we need to determine whether negotiation is an efficient and effective coordination mechanism still remains unanswered. Can we classify different forms of negotiation into complexity classes that reflect the inherent difficulty of determining whether these forms are appropriate coordination mechanisms? What are the factors that lead to this difficulty? Can this difficulty be reduced in some way? These are some of the problems confronting the computational social theorist.

One way to make some progress on these issues is to vary the properties of the negotiation model, and explore the extent to which changes in these properties affect the computational complexity of determining whether negotiation leads to an agreement [Papadimitriou, 1994]. To this end, we introduce negotiation *protocols* that define the "rules of encounter" adopted by the negotiating agents [Rosenschein and Zlotkin, 1994]. By specifying the proposals that agents are allowed to make as a function of prior negotiation history, protocols affect the degree of "sophistication" of the negotiation model. One requirement of protocols is that the number of proposals they allow at each stage of the negotiation process should be at least polynomial in the size of the negotiation scenario. This seems to be a reasonable requirement, since exponentially long series of proposals could not be enumerated in practice, and therefore protocols could never be implemented in any realistic domain. Finally, protocols are represented as a two-tape Turing machine that takes as input a representation of prior negotiation process on its first tape, and writes as output the set of possible subsequent proposals on the second tape. We further assume that the Turing machine requires time polynomial in the size of the negotiation scenario to carry out this computation.

We now consider the problem of whether a particular protocol guarantees that agreement among negotiation participants will be reached: we refer to this as the *negotiation problem*. A protocol is successful if every negotiation process compatible with the protocol ends with agreement being reached. Successful protocols are desirable, for obvious reasons. However, determining whether or not a protocol is successful is not always an easy task. In general, a protocol allows *branching* during the negotiation process. The reason for this branching is related to the fact that negotiation participants are allowed to make a number of proposals at each round. It is straightforward to see that the number of negotiation processes of length l that can be generated using a negotiation protocol with branching factor b will be b^l, that is, exponential in the length of the protocol. As a result, the problem of determining whether or not a given protocol brings about an agreement can be intuitively decomposed into an exponential number of individual computational problems. This suggests that the negotiation problem is likely to be computationally hard. The objective of our analysis is precisely to investigate how hard this problem actually is. We present a series of progressively more sophisticated protocols and we establish exactly when complexity arises. Ultimately, our aim is to identify the conditions that need to be placed upon the protocol in order to reduce the complexity of the negotiation problem and to ensure that negotiation eventually terminates with an agreement.

Conclusion

This work provides guidance for how to use logical formalisms to reason about agents' behavior and cognition during negotiation in multi-agent environments. In so doing, we have been motivated by two objectives. Firstly, to place the study of negotiation on a more secure and formal footing. In developing our formalization, we brought some of the major research questions in social sciences to bear on the methods and analytical tools advocated by mainstream computational theory. For example, we attempted to formalize such problems as the agent's bounded rationality and opportunistic behavior using a computational BDI logic. Our second objective was to identify and address an important computational problem in the use of logic-based languages for negotiation - the problem of determining whether a particular negotiation protocol will lead to an agreement. This problem is computationally hard, and the main contribution of this paper was to see just how hard that problem actually is. Obvious future lines of work are to consider the impact of these results on the design of negotiation languages and protocols, and to extend the work to cover more expressive languages for formalizing more sophisticated inter-agent dialogues.

References

[Carley and Newell, 1994] Carley, K. M. & A. Newell, 1994, "The nature of the social agent". *Journal of Mathematical Sociology*, 19(4): 221-262.

[Emerson, 1990] Emerson, E. A., 1990, "Temporal and modal logic". In van Leeuwen, J. (ed.), *Handbook of Theoretical Computer Science*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 996-1072.

[Fagin *et al*., 1995] Fagin, R., J. Y. Halpern, Y. Moses & M. Y. Vardi, 1995, *Reasoning about Knowledge*, The MIT Press, Cambridge, MA.

[Hintikka, 1962] Hintikka, J., 1962, *Knowledge and Belief*, Cornell University Press, Ithaca, NY.

[Lewicki *et al.*, 2003] Lewicki, R. J., B. Barry, D. M. Saunders & J. W. Minton, 2003, *Negotiation*, McGraw-Hill, New York, NY.

[Panzarasa *et al.*, 2001a] Panzarasa, P., N.R. Jennings & T.J. Norman, 2001, "Formalising collaborative decision-making and practical reasoning in multi-agent systems". *Journal of Logic and Computation*, 11(6): 1-63.

[Panzarasa *et al*., 2001b] Panzarasa, P., T. J. Norman & N. R. Jennings, 2001, "Social Mental Shaping: Modelling the Impact of Sociality on Autonomous Agents' Mental States". *Computational Intelligence*, 17(4): 738-782.

[Papadimitriou, 1994] Papadimitriou, C. H., 1994, *Computational Complexity*, Addison-Wesley, Reading, MA.

[Rosenschein and Zlotkin, 1994] Rosenschein, J. S. & G. Zlotkin, 1994, *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*, The MIT Press, Cambridge, MA.

[Rao and Georgeff, 1998] Rao, S. & M. P. Georgeff, 1998, "Decision procedures for BDI logics". *Journal of Logic and Computation*, 8(3): 293-342.

[Simon, 1976] Simon, H. A., 1976, *Administrative Behavior*, 3rd Edition, Free Press, New York, NY.

[Williamson, 1964] Williamson, O. E., 1964, *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*, Prentice-Hall, Englewood, Cliffs, NJ.

[Wooldridge, 2000] Wooldridge, M., 2000, *Reasoning About Rational Agents*, The MIT Press, Cambridge, MA.