BDI and BOID Argumentation

Guido Boella
Dipartimento di Informatica
Università di Torino
Italy
guido@di.unito.it

Leendert van der Torre
SEN3
CWI Amsterdam
The Netherlands
torre@cwi.nl

Abstract

In this discussion paper we are interested in the role of argumentation in the context of cognitive BDI and BOID agents, i.e., agents whose deliberation is based on beliefs, obligations, intentions and desires. We discuss argumentation issues for single agent deliberation, multiagent dialogues, and interaction between agents and their normative system. For each category we discuss examples and we give a personal view on their formalization.

1 Introduction

Argumentation has been a popular approach to non-monotonic logic since the work of Pollock, Loui and others in the eighties, who showed that argumentation is a very natural way of conceptualizing non-monotonic reasoning. In the early nineties Dung and others showed that argumentation is also very suitable as a general framework for relating non-monotonic logics of different styles. Finally, in recent years argument-based logics have been used to formalize informal theories of argumentation.

Argumentation can be studied on its own, but it also has interesting relations with other topics, such as dialogue and decision. For instance, argumentation is an essential component of such phenomena as fact finding investigations, negotiation, legal procedure and online dispute mediation. However, only recently researchers have begun to explore the use of argumentation in these contexts. We therefore organized a special session of last year's Workshop on Non-Monotonic Reasoning (NMR'2002) that aimed to bring these researchers together, to promote the logical study of argumentation and its connections with decision and dialogue.

One of the results of this workshop was the observation that formal argumentation has mainly focussed on deliberation about information and knowledge. The kind of applications mentioned above, however, typically involve formal argumentation in the context of cognitive BDI and BOID agents, i.e., agents whose deliberation is based on beliefs, obligations, intentions and desires. This kind of argumentation has been pioneered by [Parsons *et al.*, 1998,McBurney and Parsons, 2002].

We are interested in issues in argumentation and dialogue, which do not occur in single agent argumentation concerned only with information and knowledge. For example, is argumentation with respect to desires or goals different from argumentation with respect to beliefs (whatever this means precisely)? Many researchers (including referees of this workshop) argue that the state of the art argumentation frameworks are capable of dealing with desires and goals as well. However, recent results on plan argumentation of Amgoud [2003] (also see [Amgoud *et al.*, 2003b,Amgoud *et al.*, 2003a]) suggest that fundamental concepts of argumentation theory have to be adapted before argumentation theory can be used in an agent context.

Moreover, consider the distinction between single agent and multiagent argumentation. When two agents are arguing, making an argument not only may convince the other agent, but it may also give the other agent some useful information, as illustrated by the following example.

B: I want a new computer.

A: We cannot afford it now.

B: OK.

A: I want to go to Acapulco, since there is budget.

B: But then we can afford a new computer.

We present some examples, and some preliminary notes on their formalization. We discuss the following questions.

1. Which kinds of argumentation issues can be distinguished in BDI/BOID agents?

2. Which kinds of argumentation issues can be distinguished in dialogues among BDI/BOID agents?

3. Which kinds of argumentation issues can be distinguished in interactions between a BOID agent and its normative system?

In the notes on formalization we use rule based logics such as input/output logic [Makinson and van der Torre, 2000], and rule based architectures such as used in BDP logic [Thomason, 2000], and in the BOID architecture [Broersen *et al.*, 2002a,Broersen *et al.*, 2002b].

The layout of this discussion papers follows the three questions above. In section 2 we discuss single agent arguments. In Section 3 we discuss multiagent dialogues. In Section 4 we discuss interaction with normative systems.

2 Arguments

2.1 Examples

In this section we consider arguments provided by a BOID agent to justify its decisions in terms of its beliefs, obligations, intentions and desires. Consider the following example of reasoning with beliefs, obligations, intentions and desires:

1. You want to go to Acapulco for holidays;

2. You have to spend little money;

3. You intend to go to conference in Acapulco;

4. You believe that combining conference and holidays implies spending little.

Assume you combine your visit to Acapulco with a week of holidays, and someone asks you why you did so. Now you have to reconstruct an argument. You may tell that you want to go to Acapulco, and that was the reason why you did so. Alternatively, you may present the following argumentation.

- You have the normative goal to spend little;
- You believe that this can be achieved by combining conference and holidays;
- You already intend to go to conference in Acapulco;
- You therefore spend your holidays following IJCAI.

You derive from the normative goal to spend little, that you will combine conference and holidays. Note that this inference is not by application of a belief rule, only by application of the inverse of a belief rule. Consider the following alternative of the fourth sentence::

4. If you spend little, then you combine conference and holidays.

The distinction between the two sentences is familiar. The latter sentence encodes the information as a particular kind of rules used in planning, in the context of deliberation and cognitive agents such kind of rules are called practical reasoning rules. The original formalization encode the sentence as a standard belief rule. Now the intended conclusion can be achieved by means-end reasoning, i.e., planning based on abduction, as it has been studied since decades. For further discussion and examples, see [Thomason, 2000].

The aspect of this example in which we are interested, is that the abduction to find a plan can be combined with deduction to find a goal. For example, assume that the second line of the example is replaced by the following sentences.

2. If the budget is nearly finished, then you have the normative goal to spend little; You believe the budget is nearly finished;

The example can also be extended in various other ways. For example, when the third line is replaced by the following sentence, then the logic should still imply that you will submit a paper.

3. You intend to go to Acapulco if you submit a paper and this paper is accepted;

Submitting a paper does not imply that you will go to Acapulco, but it is a necessary precondition to complete the above argument.

2.2 Notes on formalization

To formalize the various kinds of rules we must distinguish between the various mental attitudes. For the extension of the example with the submission of a paper, we need decision variables and a way to deal with uncertainty, because submitting a paper does not necessarily imply that you go there. A decision theory with these ingredients have been proposed in [Dastani and van der Torre, 2002] and an argumentation theory has been proposed in [Amgoud, 2003].

The rules involved in the example above are represented by $R(l_1 \wedge \ldots \wedge l_n \rightarrow l)$, which means that the set R contains the rule having as preconditions $l_1 \wedge \ldots \wedge l_n$ and l as consequence. The set of rules involved in the example are the beliefs B, the desires D, the intentions I and obligations O:

1. $D(\top \rightarrow va)$ You want to go to Acapulco for holidays;

2. $O(\top \rightarrow \neg s)$ You am obliged to spend little money;

3. $I(\top \rightarrow ca)$ You intend to go to conference in Acapulco;

4. $B(ca \wedge va \rightarrow \neg s)$ You believe that combining conference and holidays implies spending little.

As suggested by for example the BDP logic [Thomason, 2000], the BOID architecture [Broersen *et al.*, 2002a,Broersen *et al.*, 2002b] and in our previous research [Boella and van der Torre, 2003a], we do not have to introduce modal logic to formalize such examples. However, we should distinguish between the logics of the various attitudes. For example, with rule set $R = \{c \rightarrow p\}$ we have argument for p in context c, regardless whether R stands for beliefs or obligations, but not always an argument for $c \wedge p$. The latter may make sense for beliefs, but definitely not for obligations, desires and intentions. This reflects the way belief rules and motivational attitude rules are used by BOID agents. Beliefs rules are iteratively applied to compute consequences of actions from initial states. In contrast, desire and goal rules are used to value states by checking which rules are applicable in a given state but not consistent with it.

Most logics for reasoning with or about rules do not distinguish between different logics for different kinds of rules. An exception is input/output logic [Makinson and van der Torre, 2000], which studies for example logics with or without identity, with or without transitivity, and with or without reasoning by cases. Moreover, extensions distinguish between two kinds of constraints, and notions of permissions.

A question of particular interest is how permissions are related to for example disbeliefs and undercutters, and how rule-based BOID agents can be extended to deal with them. Some preliminary observations in the context of the input/output logic framework can be found in [Boella and van der Torre, 2003c].

For further information on the BOID project, see the BOID home page:

http://boid.info/

The formalization of some of the ideas presented in this section is the subject of a cooperation between the BOID project and the IRIT laboratory in Toulouse, see:

http://boid.info/gogh

3 Dialogues

3.1 Examples

In this section we consider the case of dialogues between two agents who are arguing on some topic in order to take a decision. The computer example in the introduction shows how the information about desires, intentions and obligations which an agent puts forward can be reused in the other agent's counterarguments: if A had not mentioned that he wants to go to Acapulco, B would not have had a good argument to get his computer. A new argument not only may change the beliefs of the adversary, but it may also lead the other agent to reconsider its intentions.

Note that B's counterargument implicitly presupposes that going to the conference and buying a computer cannot be done at the same time and that buying computers is more important than going to conference. This reflects the existence of a mechanism in a BOID agent to resolve conflicts between incompatible rules. Moreover, note that A's argument in favor of going to the IJCAI'03 conference uses again the inverse of a belief rule in means end reasoning, because going to Acapulco is the means to spend the budget. Moreover, the goal which is achieved by going to the conference is justified by the obligation to finish spending the budget: agent A adopts the content of the obligations as its goal.

A variant of the example in section 2 illustrates the well known issue of intention reconsideration. B tries to change A's intention by suggesting another option he did not consider:

A: I intend to go to Acapulco for IJCAI 03 since I want holidays without spending too much;

B: There is also a conference in Kazakhstan and going to Kazakistan is less expensive than going to Mexico.

Moreover, also desires can be influenced, for example by providing new evidence:

A: I want to smoke

B: Smoking is not healthy

A: I know that

B: Here you see some lung pictures

A: I lost my appetite

Since desires have a conditional character they can be activated by making their conditions true. In this example B tries to activate the desire to stay alive by recalling A that smoking is not healthy. Unluckily this desire has been already been considered by A in its decision to smoke. In order to make A change its decision is necessary to enable some other desire in favor of not smoking. Since people usually fear to die when they realize how painful it can be, B shows A some medical picture of lung diseases.

Similar arguments to influence behavior are possible when A has a dilemma. For example, consider an agent in dilemma whether to stay or to go. For example, he may have conflicting desires, or conflicting candidate goals. In such cases, mentioning arguments to stay may have the consequence that he stays, and mentioning arguments why he should go could have the consequence that he goes.

3.2 Notes on formalization

The first example illustrates that communication can be strategic: by making an argument you also inform the other which desires, intentions and obligation you have. Reasoning about such arguments should be integrated with decision and game theory. The example can be modelled by two private knowledge bases, shared knowledge bases (issues under consideration), agent profiles, et cetera.

The second example of intention reconsideration is related to cooperation to taking a decision in a group, since it is implicit that there is a group travel budget. New information by one party can modify the decision of the group. This example presupposes some form of conflict resolution in order to take an alternative among different alternative incompatible solutions, such as the one described in [Broersen *et al.*, 2002a]).

The crucial line in the smoking example is $D_A(h \to \neg k)$: horrible means fear to die. The desire of agent A is a conditional one: unless its precondition is not true, the desire cannot be counted among the satisfied nor among the unsatisfied ones. Hence, in order to make A take this desire into account in its decision whether to smoke B must decide to show some pictures which recall A how bad is falling ill.

The remaining examples illustrate manipulation by giving the other agent another option. There are two possible ways: the agent can only give another option, and trust the decision making of the other agent to reconsider, or the agent does not trust the decision making of the other agent, and explains to him in detail which alternative is better. A reason for the first option is that if the other agent finds out himself, he thinks it is his own idea and will do it. A reason for the second option is that agents may not succeed in finding the best alternative.

4 Normative dialogues

4.1 Examples

Arguments and dialogues in a normative setting have been discussed in artificial intelligence and law, see, e.g., [Prakken and Sartor, 1996,Boella and van der Torre, 2003b]. In this section we consider dialogues involving sanction based obligations. Such sanctions are used to motivate selfish agents to respect norms.

A: I want to smoke.

N: If you smoke you violate an obligation.

A: I do not care.

N: But you got a fine of 100 euro.

A: You are too busy to apply the sanction.

In this example agent N tries to make A reconsider his intention to smoke by recalling him, first, that there is an obligation not to smoke; second, that the violation of the obligation is punished. The first argument is rejected by A since it is not a respectful agent, the second one is rejected on the ground that who is in charge of punishing violations at the moment is not able to do that.

4.2 Notes on formalization

We have studied the relation between an agent and its normative system under the assumption that the agent attributes mental attitudes to the normative system [Boella and van der Torre, 2003a,Boella and van der Torre, 2003d] and thus treats the normative system as just another autonomous agent [Boella and Lesmo, 2002]. This perspective has been useful to define the interaction between the agent and its normative system as a game between the two, such that standard game theoretic machinery can be applied. In particular, it has been useful to formalize decision-making in the context of various notions of fraud and deception. In a legal model, the normative system contains various normative agents such as legislators, judges and policemen, which can each be convinced, deceived, or bribed. In the example above, agent N is a policeman.

We believe that the normative system as agent metaphor can be used too to formalize the interaction between an agent and its normative system as a dialogue between two agents. For example, standard FIPA communication can be used to formalize the dialogue between the agent and its normative system. In order to take a decision, the agent A who is subject to the obligation has to consider the reaction he expects the normative agent N will have. This reaction is computed by recursively model N's decision using the beliefs, desires, intentions attributed to N. Finally, to formalize the example, an agent cannot be presumed to comply with obligations. In other words, it cannot be assumed that it is respectful, in the sense that it may not do what it is obliged to do. Moreover, sanctions cannot be formalized as mere consequences of violations. Instead, sanctions may be formalized as actions of the normative system, whose reaction must be taken into account in the discussion.

5 Conclusion

In this paper we illustrate that argumentation in the context of BDI and BOID agents raises new issues. First of all an argument of a BOID agent may involve reference not only to its beliefs but also to its desires, intentions and obligations. Second, in disputes between BOID agents desires, intentions and goals of both agents can be used as pros and cons. Moreover, agents try not only to make the adversary change its beliefs but they can try to make him reconsider its intentions. Third, when we consider sanction-based obligations, the agents must take into account the beliefs, goals and intentions of the normative agent who is in charge of monitoring and sanctioning violations.

Further issues to be addressed are how agents resolve conflicts among their mental attitudes when they make their decisions. Moreover, the scenario involving normative dialogues becomes more complex when we consider hierarchical normative systems composed of agents playing different roles [Boella and van der Torre, 2003b].

Acknowledgement

References

[Amgoud *et al.*, 2003a] L. Amgoud, N. Maudet, and S. Parsons. Arguments, dialogues and negotiation. In *Proceedings of ECAI'00*, 2003.

[Amgoud *et al.*, 2003b] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of ICMAS'00*, 2003.

[Amgoud, 2003] L. Amgoud. A formal framework for handling conflicting desires. In *Proceedings of ECSQARU'03*, LNAI, 2003.

[Boella and Lesmo, 2002] G. Boella and L. Lesmo. A game theoretic approach to norms. *Cognitive Science Quarterly*, 2(3-4):492–512, 2002.

[Boella and van der Torre, 2003a] G. Boella and L. van der Torre. Attributing mental attitudes to normative systems. In *Proceedings of AAMAS'03*, 2003.

[Boella and van der Torre, 2003b] G. Boella and L. van der Torre. Obligations and permissions in hierarchical normative systems. In *Proceedings of ICAIL'03*, 2003.

[Boella and van der Torre, 2003c] G. Boella and L. van der Torre. Permissions and rebutters. In *Proceedings of IJCAI'03 workshop NRAC'03*, 2003.

[Boella and van der Torre, 2003d] G. Boella and L. van der Torre. Rational norm creation: Attributing mental attitudes to normative systems, part 2. In *Proceedings of ICAIL'03*, 2003.

[Broersen *et al.*, 2002a] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

[Broersen *et al.*, 2002b] J. Broersen, M. Dastani, and L. van der Torre. Realistic desires. *Journal of applied non-classical logics*, 12(2):287–308, 2002.

[Dastani and van der Torre, 2002] M. Dastani and L. van der Torre. What is a normative goal? In *Proceedings of RASTA'02*, Bologna, 2002.

[Makinson and van der Torre, 2000] D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

[McBurney and Parsons, 2002] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.

[Parsons *et al.*, 1998] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[Prakken and Sartor, 1996] H. Prakken and G. Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4(3-4):331–368, 1996.

[Thomason, 2000] R. Thomason. Desires and defaults: A framework for planning with inferred goals. In *Proceedings of KR'00*, pages 702–713, 2000.