Preferences and Assumption-Based Argumentation for Conflict-Free Normative Agents

Dorian Gaertner and Francesca Toni

Department of Computing
Imperial College London, UK
{dg00,ft}@doc.ic.ac.uk

Abstract. Argumentation can serve as an effective computational tool and as a useful abstraction for various agent activities and in particular for agent reasoning. In this paper we further support this claim by mapping a form of normative BDI agents onto assumption-based argumentation. By way of this mapping we equip our agents with the capability of resolving conflicts amongst norms, beliefs, desires and intentions. This conflict resolution is achieved by using a variety of agents' preferences, ranging from total to partial orderings over norms, beliefs, desires and intentions, to entirely dynamic preferences defined in terms of rules. We define one mapping for each preference representation. We illustrate the mappings with examples and use an existing computational tool for assumption-based argumentation, the CaSAPI system, to animate conflict resolution within our agents. Finally, we study how the different mappings relate to one another.

Keywords: norms, BDI agents, conflicts, argumentation.

1 Introduction

Normative agents, namely agents that are governed by social norms (see for example [5,7,28]), may be subject to conflicts amongst their individual desires, or beliefs, or intentions. Such conflicts can be resolved by rendering information (such as norms, beliefs, desires and intentions) defeasible and by enforcing preferences [30]. In turn, argumentation has proved to be a useful technique for reasoning with defeasible information and preferences (e.g. see [21,23,25]) when conflicts may arise.

In this paper we adopt a model for normative agents, whereby agents hold beliefs, desires and intentions, as in a conventional BDI model, but these mental attitudes are seen as contexts and the relationships amongst them are given by means of bridge rules (as in [24]). We adopt a norm representation that builds upon and extends the one given for the BDI+C agent model of [15] and refer to our agents as BDI+N agents. In this work, norms are *internalised* as bridge rules. This representation is a natural one, in that norms typically concern different mental attitudes. Bridge rules afford a specific kind of rule-based norm

I. Rahwan, S. Parsons, and C. Reed (Eds.): ArgMAS 2007, LNAI 4946, pp. 94–113, 2008.

representation that lends itself to a mapping onto argumentation frameworks, as we show in this paper.

Furthermore, we assume that preferences over bridge rules and mental attitudes are explicitly given, to be used to resolve (potentially arising) conflicts. We consider three kinds of representations for preferences:

- by means of total orders over conflicting information;
- by means of partial orders over conflicting information;
- by dynamic rules that provide partial, domain-dependent definitions of preferences, e.g. as in [21,23,25].

For the detection and resolution of conflicts arising from choosing to adopt social norms, and for each form of preference representation, we use a specific form of argumentation, known as assumption-based argumentation [4,10,12,17,23]. This has been proven to be a powerful mechanism to understand commonalities and differences amongst many existing frameworks for non-monotonic reasoning [4], for legal reasoning [23], for practical and epistemic reasoning [17], for service selection and composition [32] and for defeasible reasoning [31]. Whereas abstract argumentation [9] focuses on arguments seen as primitive and atomic and attacks as generic relations between arguments, assumption-based argumentation sees arguments as deductions from "assumptions" in an underlying "deductive system" and defines attacks against arguments as deductions for the "contrary" of assumptions supporting those arguments.

Assumption-based argumentation frameworks can be coupled with a number of different semantics, all defined in dialectical terms and borrowed from abstract argumentation, some credulous and some sceptical, of various degrees. Different computational mechanisms can be defined to match the semantics, defined in terms of dialectical proof procedures, in particular, GB-dispute derivations [11] (computing the sceptical "grounded" semantics), AB-dispute derivations [10,11] (computing the credulous "admissible" semantics) and IB-dispute derivations [11,12] (computing the sceptical "ideal" semantics). All these procedures have been implemented within the CaSAPI system [17].

In this paper we provide a mapping from BDI+N agents onto assumption-based argumentation, and make use of the CaSAPI system to animate the agents and provide conflict-free beliefs, desires and intentions, upon which the commitments of the agents are based. The different procedures that CaSAPI implements provide a useful means to characterise different approaches that BDI+N agents may want to adopt in order to build these commitment stores.

The paper is organised as follows. Section 2 gives some background for and a preliminary definition of our BDI+N agents, focusing on the representation of norms. Section 3 gives some background on the form of argumentation we adopt and show how it can be used to detect and avoid conflicts. Section 4 presents our approach to modelling the agents' preferences (in terms of total orderings, partial orderings and dynamic preference definitions) and using these preferences to resolve conflicts in the assumption-based argumentation counterparts of BDI+N agents. Section 5 presents some formal correspondence results between the three translations. Finally, Section 6 discusses related and future work and concludes.

This paper is a revised and extended version of our previous work in [16]. In particular, Section 2 has been restructured and Section 5 has been added.

2 BDI+N Agents: Preliminaries

In this section we briefly present the notion of BDI+N agent, discuss how norms can be represented for such agents and how they are internalised. We also present an example of a normative conflict for such agents.

2.1 Background

Our BDI+N agents are an adaptation and extension of the agent model in [15], which in turn builds upon the work in [24]. The agent model of [15] adapts an architecture based on multi-context systems that have first been proposed by Giunchiglia and Serafini in [19]. Individual theoretical components of an agent are modelled as separate *contexts*, each of which contains a set of statements in a language L_i together with the axioms A_i and inference rules Δ_i of a (modal) logic. A context i is hence a triple of the form: $\langle L_i, A_i, \Delta_i \rangle$. Not only can sentences be deduced in each context using the deduction machinery of the associated logic, but these contexts are also inter-related via *bridge rules* that allow the deduction of a sentence in one context based on the presence of certain sentences in other, linked contexts.

An agent is then defined as a set of context indices $\mathcal{I}$, a function that maps these indices to contexts, another function that maps these indices to theories T_i (providing the initial set of formulae in each context), together with a set of bridge rules BR, namely rules of inference which relate formulae in different contexts. Thus, an agent can be given as follows:

$$Agent = \langle \mathcal{I},\ \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle,\ \mathcal{I} \rightarrow T_i,\ BR \rangle$$

The normative agents we are investigating are all extensions of the well-known BDI architecture of Rao and Georgeff [27] and hence the set of context indices $\mathcal{I}$ is {B, D, I}. Bridge rules are inference rules that may be ground, non-ground, or partially instantiated axioms or norm schemata.

2.2 BDI+N Agents

For BDI+N agents, bridge rules have the following syntax:

$$\begin{aligned}
BridgeRule &::= \frac{\varphi}{\psi} \\
\varphi &::= SeqLiterals \\
SeqLiterals &::= MLiteral \mid MLiteral, SeqLiterals \\
\psi &::= MLiteral \\
MLiteral &::= MentalAtom \mid \neg MentalAtom \\
MentalAtom &::= \mathtt{B}(stateterm) \\
&\quad \mid \mathtt{B}(Eitherterm \rightarrow Eitherterm) \\
&\quad \mid \mathtt{D}(Eitherterm) \mid \mathtt{I}(actionterm) \\
Eitherterm &::= actionterm \mid stateterm
\end{aligned}$$

and norms are internalised simply as bridge rules, independently of how they are represented in their corresponding norm representation language:

$$Norm ::= BridgeRule$$

Note that we distinguish between two kinds of terms: actions that an agent can execute are called *action terms*; properties that cannot be executed are called *state terms*. State terms can be brought about by executing actions represented by action terms.

This representation of norms is an adaptation of the one proposed for the BDI+C agent model of [15]. However, in [15] norms are meant to feed into a commitment store, where commitments are associated with an agent/institution component which identifies the protagonist and the subject of a commitment. Moreover, in [15], mental atoms are simply defined as follows:

$$MentalAtom ::= \mathtt{B}(term) \mid \mathtt{D}(term) \mid \mathtt{I}(term)$$

Our distinction between action and state terms leads to a refinement of the original BNF definition for a mental atom, so that executable actions are distinguished from properties. Moreover, we allow beliefs within mental atoms to be in implicative form. We restrict intentions to only concern action terms, since, intuitively, an intention is always about some future behaviour. For example, the Bible's Commandment *"You shall not covet your neighbour's wife"* is represented in BDI+N agents as [1]:

$$\frac{B(correct(bible))}{\neg D(have(neighbours\_wife))}$$

Indeed, a man cannot intend to have his neighbour's wife: he can desire it, and this may eventually result in an intention (e.g. to leave his wife which in turn is an action). Here, both $correct(bible)$ and $have(neighbours\_wife)$ are state terms.

Simple beliefs are restricted to concern state terms, since one cannot believe an action. Implicative beliefs may have either state or action terms both as antecedent and consequent. Examples of implicative beliefs are: $\mathtt{B}(sunny \rightarrow stays\_dry(grass))$ or $\mathtt{B}(goto(mecca) \rightarrow goto(heaven))$.

Finally, note that we do not allow negative terms of either kind. So, for example, we cannot represent directly $\mathtt{B}(rainy \rightarrow \neg stays\_dry(grass))$. However, this belief can be expressed equivalently as $\mathtt{B}(raining \rightarrow not\_stays\_dry(grass))$. [2]

[1] In this paper we adopt a Prolog-like convention: ground terms and predicates begin with a lower-case letter and variables begin with an upper-case letter.

[2] The relationship between $not\_stays\_dry(X)$ and $stays\_dry(X)$ can be easily expressed in assumption-based argumentation by setting appropriate definitions of the notion of contrary, as will see later.

The bridge rule given earlier is ground. An examples of a non-ground bridge rule (also referred to as a *schema*) is:

$$\frac{B(X \rightarrow Y),\ D(Y)}{I(X)}$$

expressing that, for any X and Y, if an agent believes that $X \rightarrow Y$ and it desires Y , then the agent should intend X. An example of a partially instantiated bridge rule (schema) is:

$$\frac{B(immediately(armageddon))}{\neg D(X)}$$

namely, if one believes that Armageddon will strike immediately, then one should not desire anything. Note that the first bridge rule given earlier, as well as the bridge rule:

$$\frac{B(correct(quran))}{\neg I(goto(mecca))}$$

are intuitively norms, whereas the other example bridge rules given earlier are not. A detailed analysis of what makes a rule a norm is a complex problem beyond the scope of this paper. Here, we simply assume that agents are equipped with bridge rules including norms, and focus on dealing with conflicts that may arise amongst bridge rules/norms and theories, inference rules and axioms associated to the B, D and I mental attitudes. These conflicts may not arise when agents are created. However, agents communicate with one another (and potentially sense their environment) and by doing so update their beliefs. New beliefs can trigger a norm (possibly by instantiating a norm schema) and subsequently, a new belief, desire or intention could be adopted by the agent. This may be in conflict with existing beliefs, desires or intentions, and thus commitments may be inconsistent. Equipping BDI+N agents with preferences and argumentative abilities, provides a solution to the problem of resolving these conflicts.

2.3 Example

For illustrative purposes, throughout the remainder of this paper we use an example employing agents from the ballroom scenario described in [14]. We consider a single dancer agent at a traditional ballroom. This dancer can be represented as an agent

$$\langle \mathcal{I} = \{B, D, I\},\ \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle,\ \mathcal{I} \rightarrow T_i,\ BR \rangle$$

with BR consisting (amongst others) of the following bridge rules:

$$\frac{B(X \rightarrow Y), D(Y)}{I(X)} \quad (if\ X\ is\ an\ actionterm) \tag{1}$$

$$\frac{B(X \rightarrow Y), D(Y)}{D(X)} \quad (if\ X\ is\ a\ stateterm) \tag{2}$$

$$\frac{D(X)}{I(X)} \quad (if\ X\ is\ an\ actionterm) \tag{3}$$

and inference rules in Δ_B:

$$\frac{B(X \rightarrow Y) \wedge B(X)}{B(Y)} \quad (modus\ ponens\ for\ B) \tag{4}$$

Note, that axiom (4) corresponds to modal logic schema K for beliefs, but is not present for desires and intentions since implications can be believed but neither desired nor intended. Furthermore, we do not have positive or negative introspection (modal logic schemata 4 and 5) since we exclude nested beliefs, desires and intentions for simplicity's sake. Moreover, the bridge rules BR include also ground norms using the domain language of the ballroom. We describe a selection of these norms here:

$$\frac{B(attractive(X))}{D(danceWith(X))} \tag{5}$$

$$\frac{B(sameSex(X, self))}{\neg I(danceWith(X))} \tag{6}$$

$$\frac{B(thirsty(self))}{I(goto(bar))} \tag{7}$$

Finally, one needs to define the theories T_i of the agent, detailing his initial beliefs, desires and intentions. Our dancer in question is male, not thirsty and considers his friend and fellow dancer Bob to be attractive. Hence T_B contains $B(attractive(bob))$, $B(sameSex(bob, self))$, $B(not\_thirsty(self))$. From the first belief, norm (5) and an instance of bridge rule schema (3), one can derive that our dancer should intend to dance with Bob. However, from the second belief and norm (6) one can derive the exact opposite, namely that our dancer should not intend to dance with Bob. We believe that this inconsistency is undesirable and intend to address this problem.

3 Conflict Avoidance

In this section we provide some background on assumption-based argumentation (ABA) and show how it can be used to *avoid* conflicts, in the absence of any additional (preference) information that might help to *resolve* them.

3.1 Background

An ABA framework is a tuple $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{\ } \rangle$ where

- $(\mathcal{L}, \mathcal{R})$ is a deductive system, with a language $\mathcal{L}$ and a set $\mathcal{R}$ of inference rules,
- $\mathcal{A} \subseteq \mathcal{L}$, is referred to as the *assumption* set,
- a (total) mapping $\overline{\ }$ from $\mathcal{A}$ into $\mathcal{L}$, where $\overline{\alpha}$ is referred to as the *contrary* of α.

We will assume that the inference rules in $\mathcal{R}$ have the syntax $c_0 \leftarrow c_1, \ldots c_n$. (for $n \geq 0$), where $c_i \in \mathcal{L}$. We will represent $c \leftarrow .$ simply as $c_0.$. As in [10], we will restrict attention to *flat* ABA frameworks, such that if $c \in \mathcal{A}$, then there exists no inference rule of the form $c \leftarrow c_1, \ldots, c_n \in \mathcal{R}$ for any $n \geq 0$.

Example 1. $\mathcal{L} = \{p, a, \neg a, b, \neg b\}$, $\mathcal{R} = \{p \leftarrow a. \quad \neg a \leftarrow b. \quad \neg b \leftarrow a.\}$, $\mathcal{A} = \{a, b\}$ and $\overline{a} = \neg a$, $\overline{b} = \neg b$.

An *argument* in favour of a sentence x in $\mathcal{L}$ supported by a set of assumptions X is a backward deduction from x to X, obtained by applying backwards the rules in $\mathcal{R}$. For the simple ABA framework above, an argument in favour of p supported by $\{a\}$ may be obtained by applying $p \leftarrow a.$ backwards.

In order to determine whether a conclusion (set of sentences) is to be sanctioned, a set of assumptions needs to be identified that would provide an "acceptable" support for the conclusion, namely a "consistent" set of assumptions including a "core" support as well as assumptions that defend it. This informal definition can be formalised in many ways, using a notion of "attack" amongst sets of assumptions whereby X *attacks* Y iff there is an argument in favour of some $\overline{x}$ supported by (a subset of) X where x is in Y. In Example 1 above, $\{b\}$ attacks $\{a\}$.

Possible formalisations of "acceptable" support are: a set of assumptions is

- *admissible*, iff it does not attack itself and it counter-attacks every set of assumptions attacking it;
- *complete*, iff it is admissible and it contains all assumptions it can defend, by counter-attacking all attacks against them;
- *grounded*, iff it is minimally (wrt set inclusion) complete;
- *ideal*, iff it is admissible and contained in all maximally (wrt set inclusion) admissible sets.

These formalisations are matched by computational mechanisms [10,11,12], defined as disputes between two fictional players: a proponent and an opponent, trying to establish the acceptability of a given conclusion with respect to the chosen semantics. The three mechanisms are GB-dispute derivations, for the grounded semantics, AB-dispute derivations, for the admissible semantics, and IB-derivations, for the ideal semantics. Like the formalisations they implement, these mechanisms differ in the level of scepticism of the proponent player:

- in GB-dispute derivations the proponent is prepared to take no chance and is completely sceptical in the presence of alternatives;
- in AB-dispute derivations the proponent would adopt any alternative that is capable of counter-attacking all attacks without attacking itself;

– in IB-dispute derivations, the proponent is wary of alternatives, but is prepared to accept common ground between them.

The three procedures are implemented within the CaSAPI system for argumentation [17].

In order to employ ABA to avoid (and resolve) conflicts, one has to provide a mapping from the agent representation introduced in Section 2 onto an appropriate ABA framework and choose a suitable semantics. Given such a mapping, one can then run CaSAPI, the argumentation tool, and hence *reason on demand* about a given conclusion.

3.2 Naive Translation into Assumption-Based Argumentation

In our proposed translation, one can see all bridge rules BR, theories T_i, axioms A_i and inference rules Δ_i as inference rules in an appropriate ABA framework (given below). The language $\mathcal{L}$ holds all mental atoms that make up the norms and initial theories. The $\mathcal{R}$ component holds the bridge rules, the inference rules in all theories T_i and the axioms in all A_i. Concretely, we map each norm from the set of bridge rules BR and each element of each of the theories T_i to a fact (and hence to a rule) to an inference rule in $\mathcal{R}$.

The assumption set $\mathcal{A}$ is set to $\emptyset$ in the naive translation. Thus, a definition for $\overline{\ }$ is not required.

Therefore, a naive translation of the ballroom example in Section 2.3 into an ABA framework gives $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{\ } \rangle$ [3]:

$\mathcal{L} = L_B \cup L_D \cup L_I$

$\mathcal{A} = \emptyset$

$$\mathcal{R} = \left\{ \begin{array}{l} I(X) \leftarrow B(X \rightarrow Y), D(Y), actionterm(X). \\ D(X) \leftarrow B(X \rightarrow Y), D(Y), stateterm(X). \\ B(Y) \leftarrow B(X \rightarrow Y), B(X). \\ I(X) \leftarrow D(X). \\ D(danceWith(X)) \leftarrow B(attractive(X)). \\ \neg I(danceWith(X)) \leftarrow B(sameSex(X, self)). \\ B(attactive(bob)). \\ B(sameSex(bob, self). \\ actionterm(danceWith(X)). \\ stateterm(attractive(X)). \\ stateterm(sameSex(X, Y)). \end{array} \right\}$$

Having constructed an instance of an ABA framework in this way, one can now use the CaSAPI system [17] to determine (for any semantics supported by CaSAPI) whether a given conclusion holds, and, if so, by which arguments it is

[3] All inference rules in $\mathcal{R}$ stand semantically for the set of all their ground instances. However, note that CaSAPI can often handle variables in rules.

supported. In particular, CaSAPI would allow to support the conflicting conclusions

$$I(danceWith(bob)) \text{ and } \neg I(danceWith(bob))$$

simultaneously, under any semantics. These conslusions are supported by a trivial argument with an empty set of assumptions as support. This unwanted behaviour is due to the naivity of the translation

3.3 Avoiding Conflicts Using Assumption-Based Argumentation

The conflict between $I(danceWith(bob))$ and $\neg I(danceWith(bob))$ above can be avoided by rendering the application of the two rules supporting them mutually exclusive. This can be achieved by attaching assumptions to these rules and setting the contrary of the assumption associated to any rule to be the conclusion of the other rule. This would correspond to rendering the corresponding norms/bridge rules defeasible [31,32].

In the ballroom example, the fourth and sixth rules of the naive translation above are replaced by

$$I(X) \leftarrow D(X), \alpha(X).$$

$$\neg I(danceWith(X)) \leftarrow B(sameSex(X, self)), \beta(danceWith(X)).$$

with $\mathcal{A} = \{\alpha(t), \beta(t) | t \text{ is ground}\}$ and $\overline{\alpha(t)} = \neg I(t)$ and $\overline{\beta(t)} = I(t)$.

Within the revised argumentation framework, the conflicting conclusions $I(danceWith(bob))$ and $\neg I(danceWith(bob))$ cannot be justified simultaneously. However, adopting the admissibility semantics (implemented as AB-derivations in CaSAPI), $I(danceWith(bob))$ and $\neg I(danceWith(bob))$ can be justified separately, in a credulous manner. On the other hand, adopting the grounded or ideal semantics (and GB- or IB-derivations), neither $I(danceWith(bob))$ nor $\neg I(danceWith(bob))$ can be justified, sceptically. Thus, the conflict is avoided, but not resolved. Below, we show how to resolve conflicts in the presence of additional information, in the form of preferences over norms, elements of the theories T_i, and inference rules and axioms for the different mental attitudes.

4 Conflict Resolution Using Preferences

In this section we show how to use ABA in order to reason normatively and resolve conflicts (by means of preferences) that come about by accepting or committing to certain norms, beliefs, desires or intentions. Using these preferences, we can, for example, prioritise certain beliefs over a norm or certain norms over desires. Thus, one can think of preferences as the *normative personality* of an agent. We also need to make norms and mental atoms defeasible, by using assumptions as we have done in the earlier section. For the example in Section 2.3, an agent who values norm (3) and (5) more than norm (6) will indeed intend to

dance with Bob, whereas another agent who values social conformance, such as norm (6), higher, will not have such an intention. No agent should be allowed to both intend and not intend the same thing. Similarly, simultaneously believing and not believing or desiring and not desiring the same thing is not allowed. We will adopt the following revised agent model:

$$Agent = \langle \mathcal{I},\ \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle,\ \mathcal{I} \rightarrow T_i,\ BR,\ \mathcal{P} \rangle$$

where the new component $\mathcal{P}$ expresses the agent's preferences over norms and mental attitudes. We will consider various representations for $\mathcal{P}$ below, and provide a way to use them to resolve conflicts by means of ABA. Concretely, we start with a total ordering and a cluster-based translation for conflict-resolution. Then we add more flexibility by allowing the order to be partial. Finally, we suggest a way of defining preferences using meta-rules, e.g. as done by [21,25], and following the approach proposed in [23].

In the remainder of the paper, we will refer to an agent

$$\langle \mathcal{I},\ \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle,\ \mathcal{I} \rightarrow T_i,\ BR,\ \mathcal{P} \rangle$$

as $Agent(\mathcal{P})$, and to the ABA framework resulting from applying the naive translation to $\langle \mathcal{I},\ \mathcal{I} \rightarrow \langle L_i, A_i, \Delta_i \rangle,\ \mathcal{I} \rightarrow T_i,\ BR \rangle$ as $ABA_N = \langle \mathcal{L}_N, \mathcal{R}_N, \emptyset, \overline{\ }_N \rangle$.

4.1 Preferences as a Total Ordering

The preference information $\mathcal{P}$ can be expressed as a total function that provides a mapping from bridge rules and elements of theories/axioms/inference rules to rational numbers. For now, let us assume that $\mathcal{P}$ provides a total ordering and that the type of $\mathcal{P}$ is

$$BR \cup A_B \cup A_D \cup A_I \cup \Delta_B \cup \Delta_D \cup \Delta_I \cup T_B \cup T_D \cup T_I \rightarrow \mathbb{Q}.$$

We stipulate that lower numbers indicate a higher preference for the piece of information in question. In order to translate $Agent(\mathcal{P})$ into a form that ABA can suitably handle, we propose the following mechanism. First, we generate ABA_N. Then, all rules in $\mathcal{R}_N$ are clustered according to their conclusion. Rules in the same cluster all have the same mental atom in their conclusion literal (so that fellow cluster members have either exactly the same or exactly the opposite conclusion). Next, each cluster of rules is considered in turn. All elements of each cluster are sorted in descending order $\pi_1, \ldots, \pi_n$ by decreasing preference of their corresponding norm, belief etc. Here and in the remainder of the paper, we assume a naming convention for rules whereby π_i is the name of rule $l_i \leftarrow r_i.$, where l_1 is the literal on the left-hand side of the most important rule and r_n represents the right-hand side of the least important rule.

$$l_1 \leftarrow r_1. \qquad l_2 \leftarrow r_2. \qquad l_3 \leftarrow r_3. \qquad l_4 \leftarrow r_4. \qquad \ldots \qquad l_n \leftarrow r_n.$$

Then, we employ a trick suggested in [23,10] and add a new assumption p_i to the right-hand side of each rule:

$$l_1 \leftarrow r_1, p_1. \quad l_2 \leftarrow r_2, p_2. \quad l_3 \leftarrow r_3, p_3. \quad l_4 \leftarrow r_4, p_4. \quad \ldots \quad l_n \leftarrow r_n, p_n.$$

By introducing additional assumptions into rules we make these rules defeasible and, by appropriately defining contraries, we can render conflicts impossible. We further add rules for new terms q_i of the form:

$$\begin{array}{llllll} q_2 \leftarrow r_1 & q_3 \leftarrow r_2, p_2 & q_4 \leftarrow r_3, p_3 & \ldots & q_n \leftarrow r_{n-1}, p_{n-1} \\ & q_3 \leftarrow q_2 & q_4 \leftarrow q_3 & \ldots & q_n \leftarrow q_{n-1} \\ & & q_4 \leftarrow q_2 & \ldots & q_n \leftarrow q_{n-2} \\ & & & \ldots & \ldots \\ & & & & q_n \leftarrow q_2 \end{array}$$

Intuitively, q_{i+1} holds if π_i is "selected" (by assuming p_i) and applicable (by r_i holding). Alternatively, q_{i+1} also holds if any of the other more important rules is selected and applicable. Note that there is no definition for q_1, since, as we will see below, the first rule is not intended to be defeasible.

We can now define the contraries of each of the assumptions p_i in such a way as to allow norms with a smaller subscript (higher preference) to override norms with higher subscripts (lower preference). Concretely, by setting $\overline{p_i} = q_i$ for all $i \geq 1$, a rule π_i is only applicable if assumption p_i can be made and this is only the case if q_i cannot be shown. The only way for q_i to hold is when both r_{i-1} and p_{i-1} hold (this would also make rule π_{i-1} applicable) or any of the other more important rules is applicable. Hence π_i is only applicable if π_j is not applicable for any $j < i$. Moreover, if r_1 holds, then π_1 is always applicable, as there is no way for q_1 to hold and thus p_1 can always be assumed.

After applying this procedure to all clusters, none of the clusters of rules can give rise to conflicts and since rules in different clusters have different conclusions, there cannot be any inter-cluster conflicts either. Hence, in the case of a single cluster, the resulting ABA framework $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{\ } \rangle$ with:

$$\begin{array}{l} \mathcal{L} = \mathcal{L}_N \ \cup \ \bigcup_{i=1\ldots n}\{p_i, q_i\} \\ \mathcal{R} = \{l_i \leftarrow r_i, p_i. \mid (l_i \leftarrow r_i.) \in \mathcal{R}_N\} \ \cup \ \{q_{i+1} \leftarrow r_i, p_i. \mid (l_i \leftarrow r_i.) \in \mathcal{R}_N\} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cup \ \{q_i \leftarrow q_j \mid 1 < j < i\} \\ \mathcal{A} = \bigcup_{i=1\ldots n}\{p_i\} \\ \forall p_i \in \mathcal{A} : \overline{p_i} = q_i \end{array}$$

is conflict-free. Let us consider the ballroom example from Section 2.3 again. Assume that the most important norm is (5) - $\dfrac{B(attractive(X))}{D(danceWith(X))}$ followed by norm (6) - $\dfrac{B(sameSex(X, self))}{\neg I(danceWith(X))}$ and norm schema (4) - $\dfrac{D(X)}{I(X)}$. Assume further that the premises of both norms (5) and (6) are fulfilled, unifying X with *bob*. [4] Using norm (5) we derive $D(danceWith(bob))$. Now, only norm (6) and norm schema (4) have conflicting conclusions and are grouped together for the purpose of conflict resolution. In this example, we assumed that norm (6) is more important than norm schema (4) and hence we get a cluster:

[4] Norm schemata are instantiated at this stage.

$\neg I(danceWith(bob)) \leftarrow B(sameSex(bob, self)), p_1.$
$I(danceWith(bob)) \leftarrow D(danceWith(bob)),\ p_2.$
$q_2 \leftarrow B(sameSex(bob, self)).$

and contraries: $\overline{p_i} = q_i$.

Now the mental literal $\neg I(danceWith(X))$ will be justified, but its complementary literal will not. Note that norm (7) stating that thirsty dancers should go to the bar, does not play a part in resolving the present conflict. One may therefore argue that the requirement of having a total preference order of rules is an unnatural one. For example, one may want to be able to avoid expressing a preference between certain rules that are unrelated (i.e. concerned with different, non-conflicting conclusions).

Note further, that we are adopting the *last-link* principle [25] in using preferences for resolving conflicts, which uses the strength of the last rule used to derive the argument's claim for comparison. According to this principle, the fact that norm schema (4) is *based* on a desire derived using the most important norm is irrelevant.

Once the mapping has been formulated, reasoning with the original framework is mapped onto reasoning with an ABA framework. Alternative semantics are available (in CaSAPI) to compute whether a given claim is supported.

4.2 Preferences as a Partial Ordering

We propose a different representation for preferences if the ordering of norms, beliefs, desires and intentions is not total. We replace the function $\mathcal{P}$ with a set $\mathcal{P}$ which holds facts of the form $pref(\mu_i, \mu_j)$ that intuitively express the agent's preference for norm/belief/etc. named μ_i over the one named μ_j. Note that we assume here a naming for elements of

$$BR \cup A_B \cup A_D \cup A_I \cup \Delta_B \cup \Delta_D \cup \Delta_I \cup T_B \cup T_D \cup T_I.$$

We further stipulate that $\mathcal{P}$ contains only facts about pairs of norms, beliefs, etc whose conclusions are conflicting. We deem it unnecessary to express preferences between rules that do not conflict since they will never be part of the same cluster. We will assume that this relation *pref* is irreflexive and asymmetric. It may also be appropriate to assume that *pref* is not cyclic. The asymmetry and irreflexivity requirements can be expressed as follows [5]:

$\bot \leftarrow pref(\mu_i, \mu_j) \wedge pref(\mu_j, \mu_i) \wedge \mu_i \neq \mu_j$
$\bot \leftarrow pref(\mu_i, \mu_i)$

We define a new mapping into ABA as follows. As before, we first generate ABA_N and cluster rules in $\mathcal{R}_N$ according to their conclusion. But now elements of clusters are no longer sorted by their quantitative preference, given by the total order, but instead are considered one at a time. Moreover, each rule in $\mathcal{R}_N$

[5] We refrain in this paper from axiomatising the *pref* relation and will assume instead that $\mathcal{P}$ is given so that these requirements hold.

is implicitly assumed to have the same name as the corresponding norm, belief, desire or intention.

Within the new mapping, if for a given rule we find a conflicting rule, but there is no appropriate fact in the *pref* relation, we apply the mechanism of Section 3.3 that guarantees mutual exclusion. For example, let us consider two rules π_i and π_j in the same cluster, of the form $l_i \leftarrow r_i.$ and $l_j \leftarrow r_j.$, named μ_i and μ_j respectively, where l_i and l_j are in conflict (i.e. opposite mental literals) but neither $pref(\mu_i, \mu_j)$ nor $pref(\mu_j, \mu_i)$ belongs to $\mathcal{P}$. We follow the same mechanism as in Section 3.3, adding two assumptions to the rules, yielding:

$$l_i \leftarrow r_i, p_i. \qquad l_j \leftarrow r_j, p_j.$$

and directly setting: $\overline{p_i} = l_j$ and $\overline{p_j} = l_i$. In this way, each rule is only applicable if the other one is not.

If, however, j_n facts exist in $\mathcal{P}$ ($j_n \geq 1$) expressing the agent's preference of rules named $\mu_{j_1}, \ldots, \mu_{j_n}$ over some rule named μ_i:

$$pref(\mu_{j_1}, \mu_i), \ldots, pref(\mu_{j_n}, \mu_i)$$

where $\mu_i : l_i \leftarrow r_i.$ and $\mu_{j_1} : l' \leftarrow r_{j_1}. \ldots \mu_{j_n} : l' \leftarrow r_{j_n}.$ are such that l' is the complement of l_i, then the mechanism illustrated below is employed, ensuring that the lower priority rule is only applied in case none of the "more important" ones are applicable. The rules named μ_i, $\mu_{j_1}, \ldots \mu_{j_n}$ are rewritten as

$l_i \leftarrow r_i, p_i.$
$l' \leftarrow r_{j_1}, p_{j_1}. \quad \ldots \quad l' \leftarrow r_{j_n}, p_{j_n}.$
$q_i \leftarrow r_{j_1}, pref(\mu_{j_1}, \mu_i).$
$\ldots$
$q_i \leftarrow r_{j_n}, pref(\mu_{j_n}, \mu_i).$
$q_{j_1} \leftarrow r_i, pref(\mu_i, \mu_{j_1}).$
$\ldots$
$q_{j_n} \leftarrow r_i, pref(\mu_i, \mu_{j_n}).$

where $p_i, p_{j_1}, \ldots, p_{j_n}$ are new assumptions. Finally, we set $\overline{p_i} = q_i$ and $\overline{p_{j_1}} = q_{j_1}$, $\ldots, \overline{p_{j_n}} = q_{j_n}$, and add all facts in $\mathcal{P}$ to the set of inference rules. For a more formal definition of this mapping see [32]. The resulting ABA is conflict-free.

In order to illustrate this mapping, consider again the ballroom example, where rules are named $\mu_1, \ldots \mu_{11}$ following the order in which they are presented in Section 3.2. If $pref(\mu_6, \mu_4) \in \mathcal{P}$ then in the resulting ABA framework, a subset of the set of inference rules is:

$I(X) \leftarrow D(X), p_4(X).$
$q_4(X) \leftarrow B(sameSex(X, self)), pref(\mu_6, \mu_4).$

$\neg I(danceWith(X)) \leftarrow B(sameSex(X, self)), p_6(X).$
$q_6(X) \leftarrow D(X), pref(\mu_4, \mu_6).$

$pref(\mu_6, \mu_4).$

The first rule applies, only if $D(X)$ and $p_4(X)$ both hold. However, it is defeated by the fact that the contrary of $p_4(X)$ holds. This contrary ($q_4(X)$) is dependent on $pref(\mu_6, \mu_4)$, which is true in this example. Similarly, the rule with the conclusion $\neg I(danceWith(X))$ applies, only if both $B(sameSex(X, self))$ and $p_6(X)$ hold. In our example, this rule is not defeated, since the contrary of $p_6(X)$ cannot be shown. This contrary depends on $pref(\mu_4, \mu_6)$, which does not hold. It can hence be seen how the content of $\mathcal{P}$ influences the applicability of rules.

4.3 Defining Dynamic Preferences Via Meta-rules

The relation $\mathcal{P}$ described in the previous subsection held simple facts. One can easily extend these facts into rules [6] by adding extra conditions. As an example, one could replace the fact $pref(\mu_1, \mu_2)$ with two meta-rules one stating $pref(\mu_1, \mu_2) \leftarrow sunny$ and another one stating $pref(\mu_2, \mu_1) \leftarrow rainy$. This allows the agent to change the preference between two norms, beliefs etc depending on the weather.

The addition of conditions makes the applicability of a certain norm dependent on the fulfilment of the condition and hence allows more fine-grained control over arguments. The transformation defined in the previous subsection still applies here.

Note that one can view these meta-rules themselves as norms in the sense of "one should prefer norm 1 over norm 2 whenever the sun shines". We are currently considering another kind of conflict, that contrasts $goto(bar)$ with $danceWith(X)$ since nobody can go to the bar and be on the dance-floor at the same time. Imagine the possibility of such a conflict. Then norm (7), referring to thirsty dancers, conflicts with an instance of norm schemata (4), that refers to dance intentions. A dancer that considers himself a gentleman then prefers μ_4 over μ_7, resisting the temptation to go for a drink. A selfish dancer on the other hand prefers μ_7 over μ_4. Considering yourself as a gentleman is itself a dynamic notion, that can change once the dancer has been to the bar a few times. Considering the meta-rules for preferences themselves as norms opens up many potential future investigations that we are looking forward to conduct.

5 Theoretical Considerations

In this section we show that each of the translation mechanisms proposed in the previous section is a conservative extension of the earlier mechanism, if any. For simplicity we will always assume a single cluster of preferences.

The following result, stating that given a partial order, the tranformations given in Sections 4.2 and 4.3 are equivalent, is trivial, since the two mappings return the same outcome given a partial order:

[6] Note, that these meta-rules here only concern the *pref* predicate and should not be confused with the object-level rules that act as arguments to these preference predicates.

Theorem 1. *Consider an Agent($\mathcal{P}$) such that $\mathcal{P}$ is a partial order as in section 4.2. Let $ABA_{PO} = \langle \mathcal{L}_{PO}, \mathcal{R}_{PO}, \mathcal{A}_{PO}, \overline{\ }_{PO} \rangle$ be the ABA framework resulting from applying the transformation in Section 4.2 to Agent($\mathcal{P}$) and let $ABA_D = \langle \mathcal{L}_D, \mathcal{R}_D, \mathcal{A}_D, \overline{\ }_D \rangle$ be the ABA framework resulting from applying the transformation in Section 4.3 to Agent($\mathcal{P}$). Then, for any sentence $s \in L_B \cup L_D \cup L_I$:*

- *there is an acceptable support for s wrt ABA_{PO} iff there is an acceptable support for s wrt ABA_D*

for any notion of acceptable support given in section 3.1.

The analogous result linking the mapping for total orders and partial order, given a total order as input, is easy to prove. Below, since trivially every total order is a partial order, we will use the same symbol ($\mathcal{P}$) to stand for a total order as represented in Section 4.1 and as represented in Section 4.2. Indeed, given a total order as in Section 4.1, this can be automatically mapped onto the representation in Section 4.2, by creating an element $pref(\pi_i, \pi_j)$ for every pair of elements of the cluster such that $i < j$. For a cluster with n elements, we thus obtain $\frac{n^2-n}{2}$ facts in the *pref* predicate.

Theorem 2. *Consider an Agent($\mathcal{P}$) such that $\mathcal{P}$ is a total order as in section 4.1. Let $ABA_{TO} = \langle \mathcal{L}_{TO}, \mathcal{R}_{TO}, \mathcal{A}_{TO}, \overline{\ }_{TO} \rangle$ be the ABA framework resulting from applying the transformation in Section 4.1 to Agent($\mathcal{P}$) and let $ABA_{PO} = \langle \mathcal{L}_{PO}, \mathcal{R}_{PO}, \mathcal{A}_{PO}, \overline{\ }_{PO} \rangle$ be the ABA framework resulting from applying the transformation in Section 4.2 to Agent($\mathcal{P}$). Then, for any sentence $s \in L_B \cup L_D \cup L_I$:*

- *there is an acceptable support for s wrt ABA_{TO} iff there is an acceptable support for s wrt ABA_{PO}*

for any notion of acceptable support given in Section 3.1.

This theorem can be proven as follows. First, note that, trivially, the underlying languages of the deductive systems in the two ABAs differ only in the abducibles, their contraries, and the *pref* facts, namely:
$\mathcal{L}_{TO} - (\mathcal{A}_{TO} \cup \{x | x = \overline{a} \text{ for some } a \in \mathcal{A}_{TO}\}) =$
$\mathcal{L}_{PO} - (\mathcal{A}_{PO} \cup \{x | x = \overline{a} \text{ for some } a \in \mathcal{A}_{PO}\} \cup \mathcal{P}) =$
$L_B \cup L_D \cup L_I$.
Moreover, there is a one-to-one correspondence between assumptions in the two ABAs and contraries in the two ABAs, as follows.

Suppose we have a cluster of three conflicting rules named μ_1, μ_2 and μ_3 such that each μ_i is of the form $l_i \leftarrow r_i.$, $l_1 = l_3$ and l_2 is the complement of l_1 and l_3. Let us further assume that μ_1 is preferred to μ_2 which in turn is preferred to μ_3. This total order can be expressed in terms of the representation of Section 4.2 by the facts $pref(\mu_1, \mu_2), pref(\mu_2, \mu_3)$ and $pref(\mu_1, \mu_3)$. In ABA_{TO}, the relevant part of the $\mathcal{R}_{TO}$ component for this cluster is:

$l_1 \leftarrow r_1, p_1.$ $\quad l_2 \leftarrow r_2, p_2.$ $\quad l_3 \leftarrow r_3, p_3.$
$\quad q_2 \leftarrow r_1.$ $\quad q_3 \leftarrow r_2, p_2.$
$\quad q_3 \leftarrow q_2.$

The corresponding part of $\mathcal{R}_{PO}$ in ABA_{PO} is:

$l_1 \leftarrow r_1, p'_1.$ $\quad l_2 \leftarrow r_2, p'_2.$ $\quad l_3 \leftarrow r_3, p'_3.$
$q'_1 \leftarrow r_2, pref(\mu_2, \mu_1).$ $\quad q'_2 \leftarrow r_1, pref(\mu_1, \mu_2).$ $\quad q'_3 \leftarrow r_2, pref(\mu_2, \mu_3).$
$\quad q'_2 \leftarrow r_3, pref(\mu_3, \mu_2).$
$pref(\mu_1, \mu_2).$ $\quad pref(\mu_2, \mu_3).$ $\quad pref(\mu_1, \mu_3)$

By partially evaluating the *pref* conditions, this set of inference rules can be seen to be equivalent to

$l_1 \leftarrow r_1, p'_1.$ $\quad l_2 \leftarrow r_2, p'_2.$ $\quad l_3 \leftarrow r_3, p'_3.$
$\quad q'_2 \leftarrow r_1.$ $\quad q'_3 \leftarrow r_2.$

Clearly there is a one-to-one correspondence between each p_i in $\mathcal{A}_{TO}$ and p'_i in $\mathcal{A}_{PO}$. Furthermore, there is a one-to-one correspondence between each q_i in $\mathcal{L}_{TO}$ and q'_i in $\mathcal{L}_{PO}$.

Formally, we define two mappings α_{TO-PO} and α_{PO-TO} between the languages of the two frameworks as follows:

- let p_i, p'_i be the assumptions associated with rule named μ_i in $\mathcal{R}_{TO}$ and $\mathcal{R}_{PO}$, respectively; then:
  - $\alpha_{TO-PO}(p_i) = p'_i$
  - $\alpha_{PO-TO}(p'_i) = p_i$
- let q_i, q'_i be the contraries of assumptions p_i, p'_i associated with rule named μ_i in $\mathcal{R}_{TO}$ and $\mathcal{R}_{PO}$, respectively; then:
  - $\alpha_{TO-PO}(q_i) = q'_i$
  - $\alpha_{PO-TO}(q'_i) = q_i$

  let s be any non-assumption, non-contrary, non-preference sentences in $\mathcal{L}_{TO}$ and $\mathcal{L}_{PO}$; then $\alpha_{TO-PO}(s) = \alpha_{PO-TO}(s) = s$

This mappings can be easily extended to sets of sentences.

Lemma 1. *Given any sentence $s \in \mathcal{L}_{TO}$,*

- *there is a deduction for s wrt ABA_{TO} iff there is a deduction for $\alpha_{TO-PO}(s)$ wrt ABA_{PO}.*

Given any sentence $s \in \mathcal{L}_{PO} - \mathcal{P}$,

- *there is a deduction for s wrt ABA_{TO} iff there is a deduction for $\alpha_{PO-TO}(s)$ wrt ABA_{PO}.*

As a consequence, it is easy to see that, by definition of attack:

Lemma 2. *Given any sets of assumptions $S_1, S_2 \subseteq \mathcal{A}_{TO}$,*

- *S_1 attacks S_2 wrt ABA_{TO} iff $\alpha_{TO-PO}(S_1)$ attacks $\alpha_{TO-PO}(S_2)$ wrt ABA_{PO}.*

Given any sets of assumptions $S_1, S_2 \subseteq \mathcal{A}_{PO}$,

- *S_1 attacks S_2 wrt ABA_{PO} iff $\alpha_{PO-TO}(S_1)$ attacks $\alpha_{PO-TO}(S_2)$ wrt ABA_{TO}.*

Theorem 2 is a straightforward consequence of this lemma, since all definitions of "acceptable" support are solely defined in terms of the notion of attack.

6 Conclusions

In this paper we have proposed to use assumption-based argumentation to solve conflicts that a normative agent can encounter, arising from applying conflicting norms but also due to conflicting beliefs, desires and intentions. We employ qualitative preferences over an agent's beliefs, desires and intentions and over the norms it is subjected to in order to resolve conflicts.

We provided a translation from the agent definition to an assumption-based argumentation framework that can be executed using a working prototype implementation of the query-oriented argumentation system CaSAPI. After manually applying the translation described in this paper (from the contexts, theories and preferences of a normative BDI+N agent to an argumentation framework $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{\ } \rangle$), one can execute CaSAPI and obtain a defence set containing all assumptions employed in the argument for a given claim. From these, one can derive which rules (norms or mental atoms) have been relied upon during the argumentation process. It would be useful to embed the implementation of this translation into the CaSAPI system or develop a wrapper that does the translation and employs CaSAPI.

We have considered three different notions of preference with different degrees of flexibility and expressiveness. Some theoretical considerations allowed us to show how these notions are related. Notice how our preference model (that ranks individual rules and mental attitudes) is different from the one chosen by Amgoud and Cayrol in [2], who have a preference relation over arguments such that an attack between arguments is only relevant if the attackee is not preferred to the attacker. A related approach, based on Bench-Capon's value-based argumentation framework [3] is that of Dunne et al. who developed a preference model which takes audiences into account (see [8] and [13]).

Normative conflicts have previously been addressed from a legal reasoning perspective by Sartor [30] and from a practical reasoning point of view by Kollingbaum and Norman [22]. It is traditional in the legal domain to order laws hierarchically, using criteria such as source, chronology and speciality. One such system by Garcia-Camino et al. [18] employs these criteria and a meta-order over them to solve conflicts in compound activities. As far as we know, argumentation and in particular assumption-based argumentation, has received little attention in the agent community with respect to normative conflicts.

Argumentation-based negotiation (see for example [26]) is a field of artificial intelligence that concerns itself with resolving conflicts in a multi-agent society. However, to the best of our knowledge it has hardly been used to resolve normative conflicts of the kind we study in this paper. To the best of our knowledge, the only architecture for individual agents that uses argumentation is the KGP model [20] that follows the approach of [21] to support its control component and its goal decision capability. The KGP model has been extended to support normative reasoning [29] but no conflict resolution amongst the outcomes of norm enforcement and beliefs is performed in this extension.

We have adopted a "last-link" approach to dealing with preferences in deriving conflicting conclusions along the lines of [25]. This principle employs of the strength of the last rule used to derive the argument's claim for comparison; other (potentially stronger) rules uses earlier in the derivation process are irrelevant for determining preferences. An alternative from the standard literature is the principle of the "weakest link" [1] which compares the minimum strength of the sentences used in each argument.

In the near future, we plan to research the effects of splitting the preference function into four separate ones for beliefs, desires, intentions and norms. One may be able to draw conclusions about the kind of normative personality an agent possesses depending on how these individual preference functions relate. Such relationships have been used quantitatively by Casali et al. [6] in their work on graded BDI agents.

Acknowledgements

This research was partially funded by the Sixth Framework IST programme of the EC, under the 035200 ARGUGRID project. The first author is partially supported by a PhD bursary from the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom. The second author has also been supported by a UK Royal Academy of Engineering/Leverhulme Trust senior fellowship.

References

1. Amgoud, L., Cayrol, C.: Inferring from inconsistency in preference-based argumentation frameworks. J. Autom. Reason. 29(2), 125–169 (2002)
2. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Annals of Mathematics and Artificial Intelligence 34(1-3), 197–215 (2002)
3. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. Journal of Logic and Computation 13(3), 429–448 (2003)
4. Bondarenko, A., Dung, P., Kowalski, R., Toni, F.: An abstract, argumentation-theoretic framework for default reasoning. Artificial Intelligence 93(1-2), 63–101 (1997)
5. Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., van der Torre, L.: The BOID architecture: Conflicts between beliefs, obligations, intentions and desires. In: Proceedings of AGENTS 2001, pp. 9–16. ACM Press, New York (2001)
6. Casali, A., Godo, L., Sierra, C.: Graded BDI models for agent architectures. In: Jantke, K.P., Lunzer, A., Spyratos, N., Tanaka, Y. (eds.) Federation over the Web. LNCS (LNAI), vol. 3847, pp. 126–143. Springer, Heidelberg (2006)
7. Dignum, F., Morley, D., Sonenberg, E., Cavendon, L.: Towards socially sophisticated BDI agents. In: Proceedings of ICMAS 2000, pp. 111–118. IEEE Computer Society, Los Alamitos (2000)
8. Doutre, S., Bench-Capon, T.J.M., Dunne, P.E.: Explaining preferences with argument positions. In: Kaelbling, L.P., Saffiotti, A. (eds.) IJCAI, pp. 1560–1561. Professional Book Center (2005)

9. Dung, P.: The acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming and n-person game. Artificial Intelligence 77 (1995)
10. Dung, P., Kowalski, R., Toni, F.: Dialectic proof procedures for assumption-based, admissible argumentation. Artificial Intelligence 170, 114–159 (2006)
11. Dung, P., Mancarella, P., Toni, F.: Computing ideal sceptical argumentation. Technical report, Imperial College London (2006)
12. Dung, P., Mancarella, P., Toni, F.: A dialectic procedure for sceptical, assumption-based argumentation. In: Proceedings of COMMA (2006)
13. Dunne, P.E., Bench-Capon, T.J.M.: Identifying audience preferences in legal and social domains. In: Proceedings of the 15th International Conference on Database and Expert Systems Applications, Zaragoza, Spain, pp. 518–527 (September 2004)
14. Gaertner, D., Clark, K., Sergot, M.: Ballroom etiquette: A case study for norm-governed multi-agent systems. In: Proceedings of the 1st International Workshop on Coordination, Organisation, Institutions and Norms (2006)
15. Gaertner, D., Noriega, P., Sierra, C.: Extending the BDI architecture with commitments. In: Proceedings of the 9th International Conference of the Catalan Association of Artificial Intelligence (2006)
16. Gaertner, D., Toni, F.: Conflict-free normative agents using assumption-based argumentation. In: Proceedings of the Fourth International Workshop on Argumentation in Multi-Agent Systems (2007)
17. Gaertner, D., Toni, F.: A credulous and sceptical argumentation system. In: Proceedings of ArgNMR (2007), `www.doc.ic.ac.uk/~dg00/casapi.html`
18. García, A., Noriega, P., Rodríguez-Aguilar, J.-A.: An Algorithm for Conflict Resolution in Regulated Compound Activities. In: ESAW workshop (2006)
19. Giunchiglia, F., Serafini, L.: Multi-language hierarchical logics or: How we can do without modal logics. Artificial Intelligence 65(1), 29–70 (1994)
20. Kakas, A., Mancarella, P., Sadri, F., Stathis, K., Toni, F.: The KGP model of agency. In: Proceedings of the European Conference on Artificial Intelligence, pp. 33–37 (August 2004)
21. Kakas, A., Moraitis, P.: Argumentation based decision making for autonomous agents. In: Proceedings of AAMAS 2003, pp. 883–890 (2003)
22. Kollingbaum, M., Norman, T.: Strategies for resolving norm conflict in practical reasoning. In: ECAI Workshop Coordination in Emergent Agent Societies (2004)
23. Kowalski, R.A., Toni, F.: Abstract argumentation. Journal of AI and Law, Special Issue on Logical Models of Argumentation 4(3-4), 275–296 (1996)
24. Parsons, S., Sierra, C., Jennings, N.: Agents that reason and negotiate by arguing. Journal of Logic and Computation 8(3), 261–292 (1998)
25. Prakken, H., Sartor, G.: Argument-based extended logic programming with defeasible priorities. Journal of Applied Non-Classical Logics 7(1), 25–75 (1997)
26. Rahwan, I., Ramchurn, S., Jennings, N., McBurney, P., Parsons, S., Sonenberg, L.: Argumentation-based negotiation. Knowledge Engineering Review (2004)
27. Rao, A.S., Georgeff, M.P.: BDI-agents: from theory to practice. In: Proceedings of the First International Conference on Multiagent Systems, San Francisco (1995)
28. Sadri, F., Stathis, K., Toni, F.: Normative KGP agents. Computational and Mathematical Organization Theory 12(2/3), 101–126 (2006)
29. Sadri, F., Stathis, K., Toni, F.: Normative kgp agents. Computational & Mathematical Organization Theory 12(2-3) (October 2006)

30. Sartor, G.: Normative conflicts in legal reasoning. Artificial Intelligence and Law 1(2-3), 209–235 (1992)
31. Toni, F.: Assumption-based argumentation for closed and consistent defeasible reasoning. In: Satoh, K., Inokuchi, A., Nagao, K., Kawamura, T. (eds.) JSAI 2007. LNCS (LNAI), vol. 4914, Springer, Heidelberg (2007)
32. Toni, F.: Assumption-based argumentation for selection and composition of services. In: Proceedings of the 8th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA VIII) (2007)