Coherence-Driven Argumentation to Norm Consensus

Sindhu Joseph
Artificial Intelligence Research Institute, IIIA
Spanish National Research Council, CSIC
Bellaterra (Barcelona), Catalonia, Spain

Henry Prakken
Department of Information and Computing Sciences, Utrecht University, and
Faculty of Law, University of Groningen
The Netherlands

ABSTRACT

In this paper coherence-based models are proposed as an alternative to logic-based BDI and argumentation models for the reasoning of normative agents. A model is provided for how two coherence-based agents can deliberate on how to regulate a domain of interest. First a deductive coherence model presented, in which the coherence values are derived from the deduction relation of an underlying logic; this makes it possible to identify the reasons for why a proposition is accepted or rejected. Then it is shown how coherence-driven agents can generate candidate norms for deliberation, after which a dialogue protocol for such deliberations is proposed. The resulting model is compared to current logic-based argumentation systems for deliberation over action.

Keywords

deductive coherence, norm deliberation, normative agents, argumentation

1. INTRODUCTION

The research reported in this paper is in the context of a coherence-based approach to the modelling of autonomous artificial agents. One of the fundamental properties that a human mind tries to preserve is its coherence. Any new information is tended to be evaluated for their coherence with the whole before accepting or rejecting. Taking this intuition to artificial systems, a coherence-based agent theory [14, 18] provides the agent with a mechanism to preserve the coherence of its cognitions. With this approach, beliefs, desires or intentions are only accepted if they belong to a coherent whole. That is, a coherence-based agent not only selects the set of actions to be performed, but also looks for the best set of goals to be pursued and beliefs to be accepted, making it a more dynamic model of cognitions.

In contrast, traditional BDI theories [17] do not have such a measure of coherence built into the theory. This means that agents lack the discriminative power to evaluate a cognition, thus making them less autonomous. Further, approaches that extend the BDI approach [6] equate decision making to a process to evaluate actions (intentions) with respect to certain fixed beliefs and goals. However, this makes it hard to prioritise goals or discover potential conflicts. In recent argument-based versions of BDI [3, 4, 2] goals can be prioritized and certain conflicts can be discovered. However, they tend to be more brittle since support and defeat relations between arguments and the acceptability of arguments cannot be a matter of degree, while sets of acceptable arguments cannot contain conflicts. On all these points a coherence approach is meant to provide more flexibility, since in reality support, attack and acceptability are often a matter of deg ree. One aim of this paper is to introduce coherence models as a more flexible alternative to logic-based argumentation models.

A dynamic model of agency is all the more necessary in normative agents where conflicts between private goals, beliefs and external norms are more frequent. A generic coherence-based framework was proposed in Joseph et al. [13], applying the coherence-based approach to normative reasoning of a single agent. They show how an agent driven by its coherence evaluations can decide to adopt norms when it is coherent to do so, and dynamically decide to violate a previously adopted norm when new beliefs makes it less coherent to comply with the norm. However, since they only treat a single agent case, they do not further explore the scenario where several agents can deliberate about norms that they wish to violate. In such cases, their deliberations might identify a new set of norms that are more coherent with the social goals of the normative agents.

In the present paper we address the latter topic by extend this research to a multi-agent setting, in which two coherence-driven agents aim to reach agreement about how to regulate a certain domain of interest. We aim to define a dialogue protocol for this situation and to model how the individual agents can behave within this protocol. In particular we address the following research questions:

> How can an agent generate candidate norms for deliberation?
>
> How can an agent deliberate to accept a norm proposed by another agent?
>
> How can two agents reach consensus to adopt or discard a norm?

This paper is organised as follows. In Sections 2 and 3 we present the coherence model of [13] and how it is used to model coherence-driven normative agents. In Section 4

we propose a dialogue system for two-agent deliberation on which norms to adopt. We illustrate our approach with an example in Section 5 and compare it with related research in Section 6. We conclude in Section 7.

2. COHERENCE FRAMEWORK

Since we consider coherence-driven agents, in this section we summarise a generic coherence framework that will allow us to build coherence-based agents. The framework introduced in the work of Joseph et al [13, 12] is based on Thagard's formulation of the theory of coherence as maximising constraint satisfaction [18]. The theory of coherence is based on the underlying assumption that pieces of information can be associated with each other, the association being either positive or negative. This framework differs from other coherence-driven approaches in extending agent theories [6, 15] as it modifies the way an agent framework is perceived by making the associations in the cognitions explicit in representation and analysis. That is, in this framework coherence is treated as a fundamental property of the mind of an agent. Further, it is generic and fully computational. In the following we briefly introduce the necessary definitions of this f ramework to understand the formulation of coherence-driven norm deliberation.

2.1 Coherence Graphs

The nodes in a coherence graph represent the pieces of information for which we want to estimate coherence. Examples of such pieces of information are propositions representing concepts, actions or mental states, both atomic and complex, graded and absolute. Edges between nodes may be associated with a strength, represented by a function ζ, which is derived from the underlying relation between the pieces of information. That is, if two pieces of information are related through an *explanation*, for instance, then the function ζ assigns a positive strength to the edge connecting those pieces of information. Thagard in his characterisation classifies coherence into different types, such as *explanatory, deductive, perceptual, conceptual, analogous* and *deliberative coherence*, depending on this underlying relation. Thus, we have different ζ functions for different types of coherence. The value of the function ζ, that is, the strength on an edge, may be negative or positive. Note that a zero strength on an edge implies that the two pieces of information are unrelated, which is equivalent to not having the edge connecting the pieces of information. Hence we only consider nonzero strength values on edges.

We will illustrate the definitions with the running example of Figure 1. The graph in the example is constructed with one of the inference rules of the propositional calculus, namely Modus Tollens: $(\alpha \rightarrow \beta), \neg\beta \vdash \neg\alpha$. As we gradually build our framework, we also add more sophistication to our coherence graph in this example.

Thus a coherence graph is defined as follows:

Definition 1. A *coherence graph* is an edge-weighted undirected graph $g = \langle V, E, \zeta \rangle$, where

1. V is a finite set of nodes representing pieces of information.
2. $E \subseteq \{\{v, w\} | v, w \in V\}$ is a finite set of edges representing the coherence or incoherence between pieces of information.

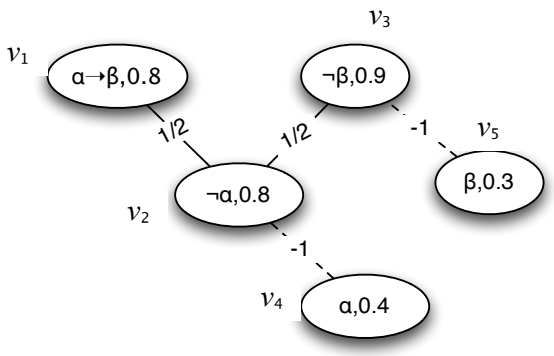

Figure 1: Graph representing the coherence and incoherence relations between graded propositions related through Modus Tollens: $(\alpha \rightarrow \beta), \neg\beta \vdash \neg\alpha$

3. $\zeta : E \rightarrow [-1, 1] \setminus \{0\}$ is an edge-weighted function that assigns a value to the coherence between pieces of information, and which we shall call a *coherence function*

Let $\mathcal{G}$ denote the set of all possible coherence graphs.

Figure 1 is an example of a coherence graph as defined above with the following values.

- $V = \{v_1, v_2, v_3, v_4, v_5\}$
- $E = \{\{v_1, v_2\}, \{v_3, v_2\}, \{v_2, v_4\}, \{v_3, v_5\}\}$
- $\zeta(\{v_1, v_2\}) = 0.5, \zeta(\{v_2, v_4\}) = -1, \ldots$

2.2 Calculating Coherence

According to coherence theory, if a piece of information is chosen as accepted (or declared true), pieces of information contradicting it are most likely rejected (or declared false) while those supporting it and getting support from it are most likely accepted (or declared true). The important problem is not to find a piece of information that gets accepted, but to know whether more than one piece of information or a set of them can be accepted together. Hence, the coherence problem is to partition the nodes of a coherence graph into two sets (accepted $\mathcal{A}$, and rejected $V \setminus \mathcal{A}$) in such a way as to maximise the satisfaction of constraints. A positive constraint between two nodes is said to be satisfied if either both nodes are in the accepted set or both are in the rejected set. Similarly, a negative constraint is satisfied if one of them is in the accepted set while the other is in the rejected set. We express this formally as follows.

Definition 2. Given a coherence graph $g = \langle V, E, \zeta \rangle$, and a partition $(\mathcal{A}, V \setminus \mathcal{A})$ of V, the *set of satisfied constraints* $C_{\mathcal{A}} \subseteq E$ is given by

$$C_{\mathcal{A}} = \left\{ \{v, w\} \in E \;\middle|\; \begin{array}{l} v \in \mathcal{A} \text{ iff } w \in \mathcal{A}, \text{ when } \zeta(\{v, w\}) > 0 \\ v \in \mathcal{A} \text{ iff } w \notin \mathcal{A}, \text{ when } \zeta(\{v, w\}) < 0 \end{array} \right\}$$

All other constraints (in $E \setminus C_{\mathcal{A}}$) are said to be *unsatisfied.*

To illustrate this, consider the partition $(\mathcal{A}_1, V \setminus \mathcal{A}_1)$ as in Figure 2. We see that, given this partition, the only satisfied constraints are those between $\{v_1, v_2\}$ and between $\{v_2, v_4\}$.

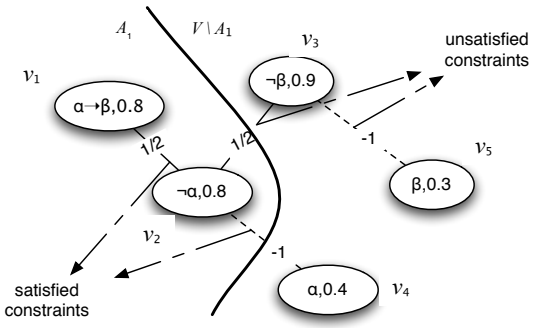

Figure 2: The strength of partition $(\mathcal{A}_1, V \setminus \mathcal{A}_1)$ is 0.375

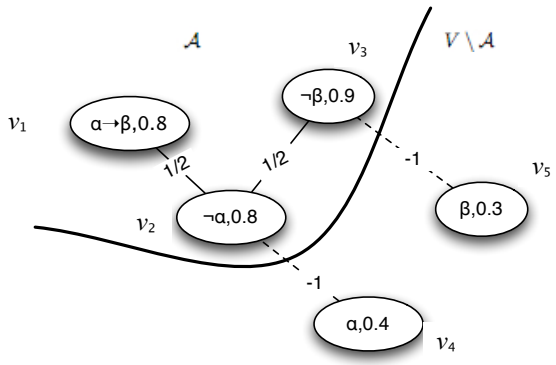

Figure 3: Coherence of the graph is 0.75 (for the partition $(\mathcal{A}, V \setminus \mathcal{A})$)

Now we define both the accepted set of the partition that maximises the satisfaction of constraints and the actual value of coherence for this partition. We first define the *strength of a partition* as the sum over the strengths of all the satisfied constraints (ζ values) of that partition. Then the coherence of a graph is defined to be the maximum among the total strengths when calculated over all its partitions. We have the following definitions:

Definition 3. Given a coherence graph $g = \langle V, E, \zeta \rangle$, *the strength of a partition* $(\mathcal{A}, V \setminus \mathcal{A})$ of V is given by

$$\sigma(g, \mathcal{A}) = \frac{\sum_{\{v,w\} \in C_{\mathcal{A}}} \mid \zeta(\{v, w\}) \mid}{\mid E \mid} \quad (1)$$

For the partition in Figure 2, the *strength* is 0.375.

Definition 4. Given a coherence graph $g = \langle V, E, \zeta \rangle$ and given the strength $\sigma(g, \mathcal{A})$ for all subsets $\mathcal{A}$ of V, *the coherence of* g is given by

$$\kappa(g) = \max_{\mathcal{A} \subseteq V} \sigma(g, \mathcal{A}) \quad (2)$$

If for some partition $(\mathcal{A}, V \setminus \mathcal{A})$ of V, the coherence is maximum, that is, $\kappa(g) = \sigma(g, \mathcal{A})$, then the set $\mathcal{A}$ is called the *accepted* set and $V \setminus \mathcal{A}$ the *rejected* set of this partition.

An important property of coherence maximisation is that the accepted set $\mathcal{A}$ is not unique. This is due to the fact that the partitions $(\mathcal{A}, V \setminus \mathcal{A})$ and its dual $(V \setminus \mathcal{A}, \mathcal{A})$ are both coherence maximising partitions. Hence, whenever $\mathcal{A}$ is a coherence maximising accepted set, so is $V \setminus \mathcal{A}$. Moreover, there could be other partitions that generate the same value for $\kappa(g)$. We state just one of the criteria to disambiguate between the accepted sets, though we are prompt to admit that there could be other criteria. If $A_1, A_2, \cdots, A_n$ are sets from all those partitions that maximise coherence of the graph g, then the coherence of the sub-graphs $(g|_{\mathcal{A}_i}, i \in [1, n])$ gives us an indication of how strongly connected they are. The higher the coherence, the better connected the pieces of information within the sub-graph. Hence, we should prefer an accepted set corresponding to the sub-gra ph with a higher coherence to that of a subgraph with a lower coherence.

For the example in Figure 1, we have a coherence maximising partition $(\mathcal{A}, V \setminus \mathcal{A})$ as in Figure 3. With this partition we see that all the constraints are satisfied and this partition gives the maximum strength for the graph.

Besides constraints, a coherence model also needs a precise definition of the coherence function ζ, i.e., of the numerical strength of coherence relations between two propositions. For reasons of space we cannot go into this here but refer the reader to [13] for the details.

2.3 Support of a node

There may be a need to find those nodes that support a given node. This is mostly required to defend a coherence-based decision. One of the criticisms raised against coherence based decision making is the lack of justification behind a decision. Here we introduce two simple notions *support set* and *conflict set* of a node, which would counter this criticism to a large extent.

The intuition is that if two nodes have a positive coherence between them, then they reinforce or give support to each other. However, we cannot take support from rejected nodes as they do not actively take part in the decision making process. And if two nodes have a negative coherence between them, then they counter each other. However, the conflict set of a node should contain the support of those nodes that conflict with the node.

Definition 5. Given a coherence graph $g = \langle V, E, \zeta \rangle$ and given a coherence maximising partition $(\mathcal{A}, V \setminus \mathcal{A})$, *the support set of a node* v is given by

$$S(v) = \{w \in \mathcal{A} | e(\{v, w\}) > 0\} \quad (3)$$

Definition 6. Given a coherence graph $g = \langle V, E, \zeta \rangle$ and given a coherence maximising partition $(\mathcal{A}, V \setminus \mathcal{A})$, *the conflict set of a node* v is given by

$$C(v) = \left\{ \begin{array}{l} w \in V | e(\{v, w\}) < 0 \\ w \in S(w) | e(\{v, w\}) < 0 \end{array} \right\} \quad (4)$$

3. COHERENCE-DRIVEN AGENTS

A *coherence-driven agent* is an agent which always takes an action based on maximisation of coherence of its cognitions, norms and other social commitments. Further, these are cognitive agents based on BDI theory [17] and are modelled as a multi-context architecture (developed by Casali et al. [7]) which consists of a set of contexts and a set of bridge rules between contexts. Each context has its own language, logic and theories expressed as coherence graphs. Bridge rules turn formulas derivable in one or more contexts into input for another context. We assume that each agent

has beliefs stored in its belief context C_B and goals stored in its desires context C_D, both individual and social goals. No relation is assumed between the C_D contexts of different agents, so they may not only have different individual goals but also have different social goals. The intentions of each agents are in the intention context C_I[1]. We also assume that each agent has a normative context C_O, which stores its opinions on the norms that should hold in the normative institution of which it is part.

3.1 Cognitive and Norm Contexts and Bridge rules

Here we briefly describe how a belief context C_B is defined while desire C_D and intentionC_I contexts are similarly defined [13, 7]. The norm logic of the norm context adapted from the work of Godo et al [9] on probabilistic deontic logic. C_B consists of a belief logic and a theory $\mathcal{T}_B$ of the logic expressed as a coherence graph.

A belief logic $\mathcal{K}_B$ consists of a belief language, a set of axioms and a deductive relation defined on the belief logic $\langle L_B, A_B, \vdash_B \rangle$. The belief language L_B is defined by extending the classical propositional language L defined upon a countable set of propositional variables PV and connectives $(\neg, \rightarrow)$. L is extended with a fuzzy unary modal operator B. The modal language L_B is built from the elementary modal formulae $B\varphi$ where φ is propositional, and truth constants r, for each rational $r \in Q \cap [0,1]$, using the connectives of Łukasiewicz many-valued logic. If φ is a proposition in L, the intended meaning of $B\varphi$ is that "φ is believable". A modal many-valued logic based on Łukasiewicz logic is used to formalise $\mathcal{K}_B$[2].

Definition 7. [7] Given a propositional language L, a belief language L_B is given by:

- If $\varphi \in L$ then $B\varphi \in L_B$
- If $r \in Q \cap [0,1]$ then $\overline{r} \in L_B$
- If $\Phi, \Psi \in L_B$ then $\Phi \rightarrow_L \Psi \in L_B$ and $\Phi \& \Psi \in L_B$ (where & and $\rightarrow_L$ correspond to the conjunction and implication of Łukasiewicz logic)

We call $\mathcal{T}_B$ a theory in the language L_B.

Other Łukasiewicz logic connectives for the modal formulae can be defined from & and $\rightarrow_L$. $\overline{0}$: $\neg_L \Phi$ defined as $\Phi \rightarrow_L \overline{0}$. Formulae of the type $\overline{r} \rightarrow_L \Psi$ (the probability of φ is at least r) will be denoted as (Ψ, r).
The axioms A_B of $\mathcal{K}_B$ are:

1. All axioms of propositional logic.
2. Axioms of Łukasiewicz logic for modal formulas (for instance, axioms of Hájek's Basic Logic (BL) [11] plus the axiom: $\neg\neg\Phi \rightarrow \Phi$.)
3. Probabilistic axioms, given $\varphi, \psi \in L$:
   - $B(\varphi \rightarrow \psi) \rightarrow_L (B\varphi \rightarrow B\psi)$
   - $B\varphi \equiv \neg_L B(\varphi \wedge \neg\psi) \rightarrow_L B(\varphi \wedge \psi)$

[1]In this paper, since we are at the level of generating obligations, we do not concentrate on actions. Actions come much later during the execution.
[2]We could use other logics as well by replacing the axioms.

The deduction rules defining $\vdash_B$ of $\mathcal{K}_B$ are:

1. Modus ponens.
2. Necessitation for B (from φ derive $B\varphi$).

Note that the truth function $\rho : L_B \rightarrow [0,1]$ is defined by means of the truth-functions of Łukasiewicz logic and the probabilistic interpretation of beliefs as follows:

- $\rho(B\varphi, r) = r$ for all $r \in Q \cap [0,1]$
- $\rho(\varphi \,\&\, \psi) = max(\rho(\varphi) + \rho(\psi) - 1, 0)$ for all $\varphi, \psi \in L_B$
- $\rho(\varphi \rightarrow_L \psi) = min(1 - \rho(\varphi) + \rho(\psi), 1)$ for all $\varphi, \psi \in L_B$

A belief graph over the belief logic $\mathcal{K}_B$ is then defined as follows:

Definition 8. Given a belief logic $\mathcal{K}_B = \langle L_B, A_B, \vdash_B \rangle$ where L_B is a belief language, A_B are a set of axioms and $\vdash_B$ are a set of deduction rules, a belief graph $g_B = \langle V_B, E_B, \zeta_B \rangle$ is a coherence graph defined over $\vdash_B$ and a finite theory $\mathcal{T}_B$ of L_B such that:

- $V_B \subseteq \mathcal{T}_B$
- E is a set of subsets of 2 elements of V_B
- ζ_B is the deductive coherence function defined over $\vdash_B$ and $\mathcal{T}_B$.

Let $\mathcal{G}_B$ denote the set of all belief coherence graphs.

A belief graph exclusively represents the graded beliefs of an agent and the associations among them. A desire graph (g_D), and a norm graph (g_O) over given logics L_D, and L_O respectively would be similarly defined. (Analogously the sets of all desire, and norm graphs are $\mathcal{G}_D$, and $\mathcal{G}_O$, respectively.)

We illustrate the concept of bridge rules with the help of an example which shows how it is used in the context of coherence graphs to reason across them.

1. Given a bridge rule $b = \frac{C_B:(B\psi,r), C_D:(D\psi,s)}{C_I:(I\psi, \min(r,s))}$ where contexts C_B, C_D, and C_I have the coherence graphs g_B, g_D and g_I associated with them respectively
2. and given $(B\psi, 0.95) \in g_B$, $(D\psi, 0.95) \in g_D$

We infer in graph g_I the intention $(I\psi, 0.95)$. Further edges with coherence values are created between $(B\psi, 0.95)$, $(D\psi, 0.95)$ and $(I\psi, 0.95)$ with coherence values equal to $\frac{2 \cdot \rho((I\psi, 0.95)) - 1}{\text{number of theory elements used in the inference}} = \frac{2 \cdot 0.95 - 1}{2} = 0.45$.

3.2 Norm Generation

We next discuss how coherence-driven agents can generate norms which, if obeyed, achieve one or more social goals that the agent thinks are important.

Conte et al [8] specify certain conditions under which an agent adopts a norm. Among other things, it has to satisfy the instrumentality condition, that the norm will be instrumental to solving some of the social or private goals of the agent. We extend the same principle to specify conditions under which a new norm is generated. A new norm we claim stems from an unsatisfied social goal and a belief that certain actions under certain conditions (can be empty) can achieve this goal. We express this with the help of a bridge

rule that says *if the goal context implies a social goal ψ and the belief context implies a belief $\phi \rightarrow \psi$ then the normative context contains an obligation ϕ.*

$$\frac{C_B : (B(\phi \rightarrow \psi), \alpha), C_D : (D\psi, \beta)}{C_O : (O\phi, f(\alpha, \beta, \gamma))}$$

If applied naively, this bridge rule will result in too many obligations: if there is more than one way to achieve ψ, then all of them will be turned into obligations, which would overconstrain the normative institution: what we want instead is to make only one way to achieve the social goal obligatory, to leave the agents' degree of autonomy as large as possible. Another aspect not taken into account by this bridge rule is that realising ϕ may frustrate another social goal, i.e., it may hold that $\phi \rightarrow \neg\psi'$ where ψ' is another social goal of the agent.

To deal with these problems, the obvious similarity can be exploited between this bridge rule and the well-known practical syllogism "If I want ψ and ϕ realises ψ, then I should intend to do ψ" . Walton (1996) formulated this as one of his presumptive argument schemes, with as main critical questions "are there other ways to realise ψ" and "does ϕ also have unwanted consequences?". In recent years several AI researchers have formalised this argument scheme in formal argumentation systems (e.g. [3, 4, 2]). The key idea here is that positive answers to Walton's two critical questions give rise to counterarguments.

Our task is to model the same idea in our coherence approach. As coherence theory is developed to make sense of such contradictions between pieces of information and identify those that cohere most together, modelling the above scenario is natural using this theory. Coherence maximisation partitions the cognitions including the obligations that selects the most coherent set of cognitions and obligations. Note that the basic relationship we model here is that between goals and norms, i.e, between goals and actions, in which different ways to achieve the same goal negatively cohere with each other. However, our framework uses only deduction as the underlying relation in which the set $\{p \rightarrow g, q \rightarrow g, p, q\}$ is consistent (here p and q are different ways to achieve goal g). Hence we add an additional explicit constraint to make these alternatives. That is, for each goal g in an agent's desire context, we consider the set of all implications $p_1 \rightarrow g, \cdots, p_n \rightarrow g$ in its belief context and we add formulas $\neg(Op_i \& Op_j)$ to its norm context for all p_i and p_j such that $1 \leq i < j \leq n$. Then two obligations Op_i and Op_j will negatively cohere with each other since they are alternatives.

The just-explained method deals with the first of Walton's critical questions of the practical syllogism (are there alternative ways to realise the same goal?). To deal with his second critical question (does ϕ also have unwanted consequences?) a bridge rule is needed that expresses the negative version of the practical syllogism: if the goal context implies a social goal ψ' and the belief context implies a belief $\phi \rightarrow \neg\psi'$ then the normative context contains an obligation $\neg\phi$.

$$\frac{C_B : (B(\phi \rightarrow \neg\psi), \alpha), C_D : (D\psi, \beta)}{C_O : (O\neg\phi, f(\alpha, \beta, \gamma))}$$

Then, in cases where an action achieves some but frustrates other social goals, our deductive coherence measures makes the obligations that result from the positive and negative version of the practical syllogism negatively cohere with each other.

4. DEFINITION OF PROTOCOL AND RELATED NOTIONS

We next present a protocol for two-agent deliberation about norm proposals. The general idea is that during a dialogue the agents jointly build a coherence graph, which is used to define turntaking and agreement. Such a joint dialogical structure distinguishes deliberation from negotiation, since unlike in negotiation, in deliberation the reasons for an agreement should be public.

More precisely, during a dialogue the agents exchange arguments, which can contain norm proposals (by applying one of the two bridge rules) or can be about goals or matters of belief. The joint coherence graph incorporates these arguments in nodes corresponding to the arguments' conclusions and premises, and in the relevant positive and negative constraints between these nodes. At each stage of the dialogue preferred partitions of the joint graph can be identified for each player, which are the partitions in which their norm proposals are best satisfied. The player with the most coherent preferr ed partition is the current winner. As soon as a player has made himself the current winner, the turn shifts to the other player, who must then try to make herself the current winner, and so on. This forces the players to make relevant moves that improve their position. A dialogue ends in agreement when both players' preferred partitions accept the same set of norms.

We now formally define the topic and communication languages L_t and L_c and the protocol. Agents choose from the dialogue moves available by incorporating each utterance from the other agent into their internal coherence graph and then choosing their reply on the basis of their internal coherence calculations. So in the end three coherence graphs are relevant: the agents' internal graphs and the joint, public one.

Let the topic language L_t consist of the union of the agents' context languages. Recall that the context languages of the agents' are the belief L_B, desire L_D, intention L_I, and norm L_O languages as defined in Section 3.1. Then L_c consists of expressions Φ *since* Γ such that Φ and all elements of Γ are well-formed formulas of L_t (below such expressions will be called *arguments*). A *move* is a pair (p, x) where x is an expression from L_c and p is the player who utters x (sometimes we will abuse notation and refer to x only as a move, leaving the speaker implicit). A *dialogue* is a sequence of moves. For any dialogue $d = m_1, \ldots, m_n, \ldots$ the sequence $m_1, \ldots, m_i$ is denoted by d_i, where d_0 denotes the empty dialogue. For any dialogue d and move m the notation d, m stands for the result of appending m to d, i.e., for d as continued by m.

Definition 9. For any dialogue d the *joint coherence graph* $g(d) = \langle V(d), E(d), \zeta(d) \rangle$ associated with d is defined as follows (we leave the coherence function implicit since it can be deduced from the other elements by the definitions of [13]):

- $V(d_0) = E(d_0) = \emptyset$ while $\zeta(d)$ is undefined;
- For any move $m = \Phi$ *since* Γ :

- $V(d,m) = V(d) \cup \{\varphi\} \cup \Gamma \cup C$, where:
  - if $m = (O\psi, \alpha)$ *since* $(B(\psi \rightarrow \chi), \beta), (D\chi, \gamma), S$ then $C = \{(\neg(O\psi \wedge O\psi'), f(\alpha, \alpha')) \mid d$ contains a move with argument $(O\psi', \alpha')$ *since* $(B(\psi' \rightarrow \chi), \beta'), (D\chi, \gamma), S$ such that $\psi \neq \psi'\}$;
  - otherwise $C = \emptyset$
- $E(d,m) = \{(v, v') \mid v, v' \in V(d,m)$ and $\zeta_(v, v')$ is defined$\}$

The joint coherence graph is initially empty. Each move adds its premises and conclusion as new nodes, after which the edges and coherence values are recalculated according to the definitions of Section 2. In addition, if a move proposes a norm in alternative to an earlier proposal for the same goal, we also add the corresponding constraint between the two norms as a new node.

Definition 10. A norm $(O\phi, \alpha)$ is *proposed* by player p in dialogue d if d contains a move (p, x) where the conclusion of x is $(O\phi, \alpha)$.

A goal $(D\psi, \gamma)$ is *addressed* by p in d if d contains a move (p, x) where x is of the form $(O\phi, \alpha)$ *since* $(B(\phi \rightarrow \psi), \beta), D\psi, S$.

A partition $(\mathcal{A}, V \setminus \mathcal{A})$ of $g(d)$ is *potentially preferred* by player p if the accepted set $\mathcal{A}$ of the partition contains a norm proposed by p for each goal addressed by p in d.

A partition $(\mathcal{A}, V \setminus \mathcal{A})$ of $g(d)$ is *preferred* by player p if it is potentially preferred by p and there is no other potentially preferred partition of $g(d)$ by p with a higher coherence value. Let $P_p(d)$ be any partition of $g(d)$ preferred by p.

Definition 11. A player p is the *current winner* of a dialogue if the coherence of its preferred partitions of $g(d)$ is higher than the coherence of the preferred partitions of $g(d)$ of its opponent. If the coherence values of both sets of preferred partitions are the same, then there is said to be no current winner.

A *protocol* Pr is a function that assigns to any legal dialogue a set of moves which are its legal continuations. A dialogue is *legal* if any move in it is a legal continuation of the sequence to which it is appended. If $Pr(d) = \emptyset$ then d is a *terminated* dialogue.

Our protocol assumes that each dialogue is against the background of a set $F = \{(D\psi_1, \alpha_1), \ldots, (D\psi_n, \alpha_n)\}$ of *focal goals*, and contains the following rules.

Definition 12. For any dialogue d, $m = (p, x) \in Pr(d)$ iff:

- p is the player to move in d;
- if $d = d_0$ then s is of the form $(O\phi, \alpha)$ *since* $(B(\phi \rightarrow \psi), \beta), (D\psi, \gamma), S$ where $(D\psi, \gamma)$ is a focal goal;
- $E(d,m)$ contains positive support links from each premise of x to its conclusion;
- if the coherence value of p's preferred partitions in $g(d,m)$ is not higher than the coherence value of p's preferred partitions in $g(d)$, then
  - either m is p's first proposal for a goal addressed in d;
  - or m repeats a proposal for a norm by p'.
- d contains no move (p, x);
- the players do not agree in d.

Furthermore, we have that player p is to move in d_i if either p' is the current winner in d_i or there is no current winner in d_i and p was to move in d_{i-1}.

To comment on these rules, the first rule is obvious while the second rule says that each discussion starts with a proposal for a norm that (if complied with) achieves some social goal. Each next move may be an argument of any form, as long as it respects the remaining protocol rules. Rule (3) says that each move must be an argument in that in the resulting joint coherence graph the premises of the move must positively cohere with its conclusion. Rule (4) says that each move must either improve the position of the speaker, or make the speaker's first norm proposal for a goal addressed in d, or accept a norm proposal by the other party. Rule (5) prevents a player from repeating his own moves, while rule (6) makes sure that a dialogue terminates after the players have reached agreement.

For defining agreement we need the following notation. For any partition $P = (\mathcal{A}, V \setminus \mathcal{A})$ of graph g let $N_p(P)$ denote the norms proposed by p belonging to $\mathcal{A}$.

Definition 13. The players p and p' *agree* in dialogue d if all focal goals have been addressed in d and there exist preferred partitions P_p and $P_{p'}$ of $g(d)$ such that $N_p(P_p) = N_{p'}(P_{p'})$.

In words, the players agree if they have discussed all focal goals and if they have preferred partitions that contain the same set of norms for all goals addressed in the dialogue (which may include more goals than just the focal goals, namely, if a move has introduced a new goal).

4.1 Internal Deliberation

We now sketch the internal deliberation of each player p to generate and evaluate proposals. Coherence-driven agents make decisions based on coherence maximisation, same is true for the cases of generation and evaluation of proposals.

4.1.1 Generate a New Move

We assume that at any time the coherence graph of an agent is closed under the application of the bridge rules. The accepted set resulting from the coherence maximising partition is the base for generating new moves. Moves are of the form Φ since Γ. Any element of the accepted set can be Φ and Γ then is the set of support nodes of Φ. Among the possible Φ's, an element is chosen based on its priority. In the case where the deliberation is on norms, norms can be given priority over other elements. Given the composite coherence graph of the agent $g = \langle V, E, \zeta \rangle$, agent performs the following to generate a new move:

1. For all partitions $(\mathcal{A}_i, V \setminus \mathcal{A}_i)$, $\mathcal{A}_i \subseteq V$ calculate the coherence $\sigma(\varsigma_{g'}, \mathcal{A}_i)$ using Equation 1.
2. Using Equation 4 from Section 2, finds a coherence maximising partition $\mathcal{A} = \mathcal{A}_i | \max(\sigma(\varsigma_{g'}, \mathcal{A}_i))$. Note that there may be more than one such partitions (preferences can be set based on discussions on Section 2).
3. $\Phi = (a\varphi, \alpha)$ such that $\alpha = \max(r | (a\varphi, r) \in \mathcal{A})$ where $a \in \{B, D, I, O\}$. (In the case of moves about norms,

$a = O$ and $\mathcal{A}$ is $V_N|_{\mathcal{A}}$ and $(O\varphi, \alpha) \notin g(d)$ (not a previously proposed norm, g(d)= joint coherence graph).

4. The support set $S(\Phi) = S(a\varphi, \alpha)$.

5. Return the dialogue move $m = \Phi$ since $S(\Phi)$

6. If $\Phi = null$, then m is set to $null$.

4.1.2 Evaluate a move

The internal deliberation of player p is similarly based on coherence maximisation. p introduces the received move into its respective coherence graphs and recalculates the composite coherence graph. Upon maximising coherence, if the elements of the move belong to the accepted set of its coherence maximising partition, it accepts the move. Else generates the reasons for rejecting a move. Given the proposed move $m = (\Phi, S(\Phi))$, a coherence-driven agent,

1. Recompute the composite coherence graph $g = \langle V, E, \zeta \rangle$ with the elements of m using bridge rules.

2. For all partitions $(\mathcal{A}_i, V \setminus \mathcal{A}_i)$, $\mathcal{A}_i \subseteq V$ calculate the coherence $\sigma(\varsigma_{g'}, \mathcal{A}_i)$ using Equation 1.

3. Using Equation 4 from Section 2, finds a coherence maximising partition $\mathcal{A} = \mathcal{A}_i | \max(\sigma(\varsigma_{g'}, \mathcal{A}_i))$.

4. If $\Phi \in \mathcal{A}$ and $S(\Phi) \subseteq \mathcal{A}$, then accept m. Else calculate the *conflict set* $C(\Psi)$ for each Ψ such that $\Psi \in \{\Phi\} \cup S(\Phi)$ and $\Psi \notin \mathcal{A}$.

5. EXAMPLE — NORM NEGOTIATION

Now we take a real scenario in which two coherence-based agents discuss certain norms for regulating a discussion forum, especially on how often the participants may reply to each others' contributions. The focal goals of the agents are:

- f = efficiency (the discussion should not take too long)
- s = coverage (the discussion should cover as much relevant material as possible)
- p = fairness (the participants should be treated fairly compared to each other)
- t = quality of contributions (the participants should be stimulated to write high-quality contributions).

In addition, one of the agents has a secret private goal $u = x$ not become a moderator.

With this background, two of the possible ways to achieve these goals and how far they help achieve each of the focal goals are given below:

1. r: *everyone gets one reply.* This promotes efficiency $(r \rightarrow f)$ and quality of individual contributions ($r \rightarrow t$) but demotes coverage $(r \rightarrow \neg s)$. The reason why this promotes quality of contributions is that with just one possible reply everyone will make it as good as possible, since they will not get a second chance. It has no net effect on fairness since on the one hand everyone gets the same number of replies (which is fair) but on the other hand an expert in the field will get less opportunity to say what he wants to say than a layman (which is unfair).

2. q: everyone may reply as long as allowed by the moderator. This also promotes efficiency $(q \rightarrow f)$ since the moderator can be trusted to keep discussions short. It also promotes fairness $(q \rightarrow p)$ since the moderator can be trusted to give experts more replies than novices. It has no particular effect on coverage or quality of contributions (since judging whether everything has been covered is too difficult for the moderator).

Hence each agent initially has the following theory. Agent

$Theory$	$\mathcal{A}$	$V \setminus \mathcal{A}$
$\mathcal{T}_N$	$(Oq, 1), (O\neg r, 0.8)$	
$\mathcal{T}_B$	$(B(q \rightarrow f), 1), (B(q \rightarrow p), 0.9)$ $(B(r \rightarrow \neg s), 1)$	
$\mathcal{T}_D$	$(Df, 1), (Dp, 0.9), (Ds, 0.8)$	

Table 1: The initial theory of Agent A

A is aware of the social goals f, p and s. Further, it knows that q helps achieve two of the goals namely f and p. Hence the initial coherence graph of the agent (Figure 4) generates norms $(Oq, 1)$ and $(O\neg r, 0.8)$. Since A so far has no incoherence, nor any other ways of achieving its goals, every element falls in the accepted set. Since Oq is preferred over $O\neg r$, A initiates the deliberation protocol with the proposal for norm $(Oq, 1)$, $d = d_0$ (the dialogue moves are in Table 4).

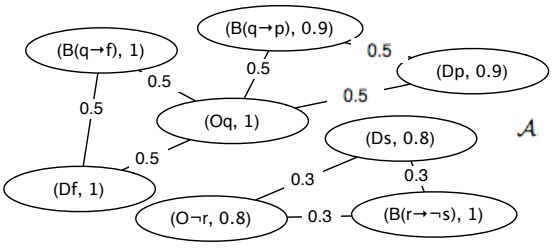

Figure 4: The initial coherence graph of Agent A

Agent B, however, has knowledge of the focal goals f and t. it also knows that r helps achieve its goals f and t. It also has a secret private goal u and it knows that $q \rightarrow \neg u$ (Table 2). Hence B generates two norms $(Or, 0.9)$ and $(O\neg q, 0.9)$. B also has all the elements in the accepted set so far (in Figure 5), as it does not yet know of the conflict between Or and Oq.

$Theory$	$\mathcal{A}$	$V \setminus \mathcal{A}$
$\mathcal{T}_N$	$(Or, 1), (O\neg q, 1)$	
$\mathcal{T}_B$	$(B(r \rightarrow f), 1), (B(r \rightarrow t), 0.9)$ $(B(q \rightarrow \neg u), 0.9)$	
$\mathcal{T}_D$	$(Df, 1), (Dt, 0.8), (Du, 0.9)$	

Table 2: The initial theory of Agent B

After A's move B updates its coherence graph with the proposed norm and its supports. However, it is natural to assume that the agents may not have the same preferences on goals. Hence, even though agent B incorporates the new information into its theory, the degrees of these cognitions

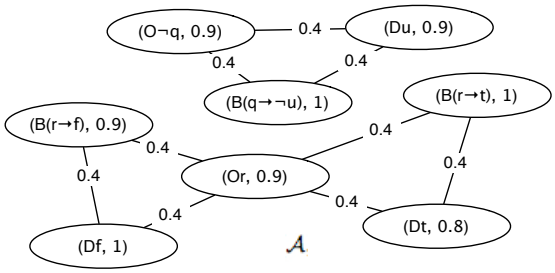

Figure 5: Initial coherence graph of B

vary according to the preference ranking of the goals. The updated theory is in Table 3 and the corresponding coherence maximising partition is in Figure 6. Since B's coherence maximising partition rejects $(Oq, 0.9)$, B makes a counter-proposal for the norm $(Or, 0.9)$.

Theory	$\mathcal{A}$	$V \setminus \mathcal{A}$
$\mathcal{T}_N$	$(Or, 1), (Oq, 1)$	
$\mathcal{T}_B$	$(B(r \rightarrow f), 0.9), (B(r \rightarrow t), 1),$ $(B(q \rightarrow f), 1), (B(q \rightarrow p), 0.9)$	
$\mathcal{T}_D$	$(Df, 1), (Dt, 0.9), (Dp, 0.8)$	

Table 3: Theory of agent B after dialogue d_0

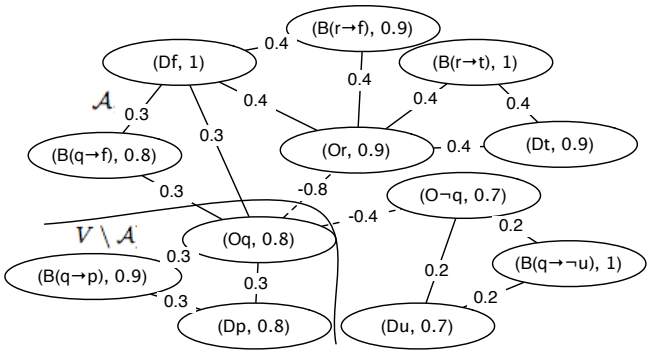

Figure 6: Coherence graph of B after dialogue d_0

Since A has no knowledge of the norm $(Or, 0.9)$ it adds the norm and the support nodes to its coherence graph. However, A finds out that r upsets its social goal s. The coherence maximisation hence rejects $(Or, 0.9)$ as in Figure 7.

B incorporates the new information about Or into its theory and calculates the new coherence maximising partition as shown in Figure 8. Due to the fact that the norm $(Or, 0.9)$ upsets social goal s in addition to the competition it has from $(Oq, 1)$, the coherence maximising partition now rejects the norm $(Or, 0.9)$ along with the private goal u, the social goal t and the beliefs relating them. Hence B proposes the only norm in its accepted set $(Oq, 0.9)$. Now the preferred partitions of both A and B in the joint coherence graph contain the single norm Oq proposed by both of them,

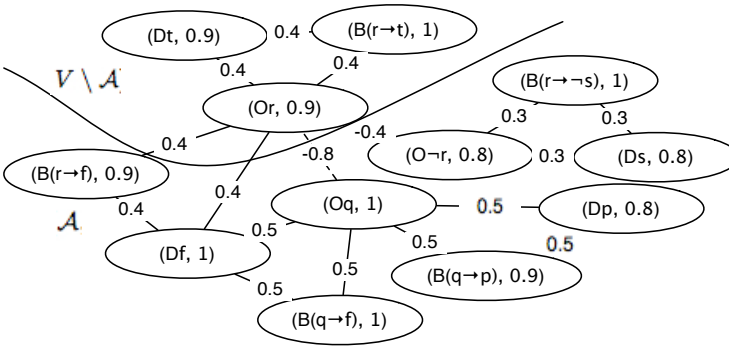

Figure 7: Coherence graph of A after dialogue d_1

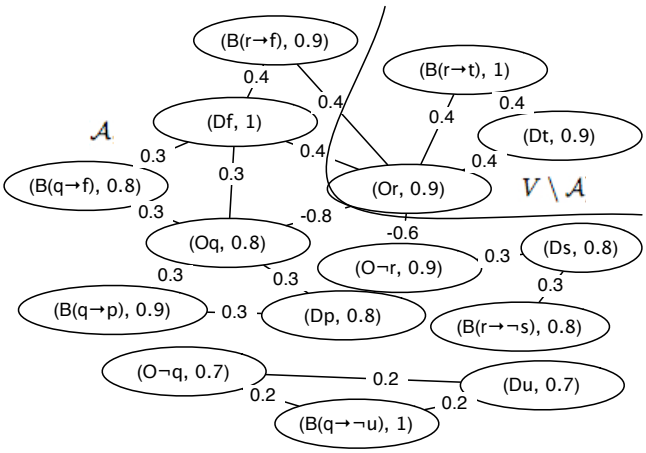

Figure 8: Coherence graph of B after dialogue d_2

so the dialogue ends in agreement. Note that due to space limitations, we could not include the evolution of the joint coherence graph, however, the final joint graph has the same nodes, the same coherence maximising partition as Figure 7, differing only in some of the degrees of the nodes. This is due to the fact that agent A has no private goals and hence reveals every information it has during the dialogue. The difference in degrees is due to the fact that , when both A and B has moves that contain common elements, the degree on the joint coherence graph is determined by taking the average of degrees. The entire dialogue is as shown in Table 4.

6. RELATED RESEARCH

Coherence models have been earlier applied to legal reasoning by Thagard [19], Amaya [1] and Bench-Capon & Sartor [5]. Thagard and Amaya use explanatory coherence to model scenario-based reasoning about evidence, while Bench-Capon & Sartor use a coherence model in their theory formation approach to case-based reasoning. Thus these proposals model different aspects than ours; moreover, they do not provide protocols for multi-agent deliberation. The most important difference however is that here we provide

Dialogue Id	*Agent*	Φ	Γ
d_0	A	$(Oq, 1)$	$\{(Df, 1), (B(q \rightarrow f), 1),$ $(Dp, 0.9), (B(q \rightarrow p), 0.9)\}$
d_1	B	$(Or, 0.9)$	$\{(B(r \rightarrow f), 0.9), (Df, 1)\}$ $(Dt, 0.9), (B(r \rightarrow t), 1)\}$
d_2	A	$(O\neg r, 0.9)$	$\{((B(r \rightarrow \neg s), 1), (Ds, 0.8)\}$
d_3	B	$(Oq, 0.9)$	$\{(Df, 1), (B(q \rightarrow f), 1),$ $(Dp, 0.9), (B(q \rightarrow p), 0.9)\}$

Table 4: Dialogues between agents A and B

a fully computational model of coherence.

We next compare our model to proposals that use logic-based argumentation. We know of no such proposals that address the problems of norm generation and normative agreement. However, norm generation is similar to intention generation by an agent who reasons how to achieve its goals, while normative agreement is similar to reaching agreement on a course of action to solve a problem. For both phenomena logic-based argumentation models have been proposed, so we will compare our model to these.

We must first distinguish between logics and protocols for argumentation. The former define which conclusions can be drawn from a given body of information, while the latter regulate how such a body of information can be constructed in dialogue. Several argument-based logics for intention generation have been proposed. Bench-Capon & Prakken [4] aim to formalise the reasoning model underlying Atkinson's [3] dialogue model for disputes over action, and by Amgoud & Prade [2]. The essential ingredient in both approaches consists of two rules for constructing arguments that correspond to our two bridge rules. Bench-Capon & Prakken then apply Prakken's [16] accrual mechanism to aggregate arguments for or against the same intentions, while Amgoud & Prade leave the aggregation of such arguments outside the logic and model it decision-theoretically.

We first note a difference in applying the first bridge rule (the positive practical syllogism), arising from the difference between intentions and norms. While [4] allow to conclude Dr from Dp, $q \Rightarrow p$ and $r \Rightarrow q$ by chaining two applications of the practical syllogism, we don't allow such chaining but only allow to conclude Oq. This is deliberate, since we want to respect the agents' autonomy to decide for themselves how they will comply with the norms they are facing.

The logics of [4, 2] instantiate the general framework of Dung [10], which starts from a set of arguments with a binary defeat relation and then determines which sets of arguments can be accepted together. This is similar to determining partitions of a coherence graph, but in approaches that instantiate Dung's format support and defeat relations between arguments and the acceptability of arguments cannot be a matter of degree, while sets of acceptable arguments cannot contain conflicts. As remarked in the introduction, on all these points a coherence approach is meant to provide more flexibility, since in reality support, attack and acceptability are often a matter of degree. In this paper we have seen one benefit of this, namely, a natural modelling of accrual of arguments for the same conclusion. By contrast, in [2] accrual is modelled outside the logic while the logical accrual mechanism of [4] is quite complex. In future research we aim to investigate whether the added flexibility of our coherence approach has other advantages.

On the other hand, a strong point of argument-based approaches is that they yield explicit reasons why an outcome should be adopted or rejected, while coherence-based approaches are often criticised for their lack of transparency. In our approach we have addressed this problem by deriving our coherence measures from the deduction relation of an underlying logic, thus making explicit why two pieces of information are positively or negatively related. This feature was then exploited in our protocol, which contains the notion of an argument.

To compare our protocol with logic-based protocols for reaching agreement over action, the most detailed proposal we know of is that of Atkinson [3], who derives a dialogue protocol from an extended version of Walton's [20] argument scheme for justifying actions and its critical questions. Let us see to what extent our protocol allows her dialogue moves to be moved as arguments in reply to an application of the first bridge rule. Let it be of the form $O\phi$[3] *since* $B(\phi \rightarrow \psi), D\psi$. Note first that we have a restricted domain ontology in that unlike Atkinson we do not distinguish between goals and values, between truth and possibility and between circumstances and actions. All these simplifications are meant to focus on the essence of our proposoal, which is its use of the coherence mechanism. These simplifications make that only a number of Atkinson's critical questions are relevant for our model (since we do not distinguish between values and goals, we have replaced Atkinson's term 'value' in CQs 9 and 10 by 'goal'):

- *CQ1: Are the believed circumstances true?* Since we model the deliberation of normgivers on how to regulate a domain instead of the reasoning of agents on what to do in concrete situations, this question is irrelevant for us.
- *CQ2: Assuming the circumstances, does the action have the stated consequences?* This can be addressed with an argument for conclusion $B(\neg(\phi \rightarrow \psi))$. This move will introduce a negative coherence link between this conclusion and the original belief $B(\phi \rightarrow \psi)$.
- *CQ5: Are there alternative ways of realising the same consequences?* This can be formulated with an alternative application of our first bridge rule: $O\phi'$ *since* $B(\phi' \rightarrow \psi), D\psi$. Combined with the constraint $\neg(O\phi \wedge O\phi')$ introduced by this move , this move adds a negative support link between $O\phi$ and $O\phi'$.
- *CQ9: Does doing the action have a side effect which demotes someother goal?* We can express this by an application of the second bridge rule. This adds a node $O\neg\phi$ to the joint coherence graph, which negatively coheres with the node $O\phi$.
- *CQ10: Does doing the action promote some other goal?* We can express this by applying the first bridge rule to the other goal, resulting in another argument for the same norm. As shown above, this normally improves the speaker's position and thus naturally models accrual of arguments.
- *CQ11: Does doing the action preclude some other action which would promote some other goal?* This corresponds to the situation that we have $B(\phi \rightarrow \neg\psi)$

[3]Here the grades are ignored for convenience.

and $B(\psi \rightarrow \chi)$ and $D\chi$. Space prevents us from going into logical detail here. Roughly, we can only express this if $\psi \rightarrow \chi$ is necessarily true, i.e., true in all possible worlds: then the argument for $O\phi$ can be countered with an argument for $O\psi$ applying the first bridge rule and further extended to $O\neg\phi$: then $O\phi$ and $O\neg\phi$ negatively cohere in the joint coherence graph.

Concluding, given our restricted domain ontology, our model essentially allows for all argument moves and critical questions proposed by Atkinson; a possible advantage of our approach over Atkinson's is a natural way to model accrual of alternative arguments for the same norm (which is arguably more natural than [4]'s logic-based model of accrual). We leave it for future research to generalise our domain ontology to the full case of Atkinson and to investigate other possible advantages of our approach over theirs.

7. CONCLUSION

In this paper we have proposed coherence-based models as an alternative to logic-based BDI and argumentation models for normative reasoning. In particular, we have provided a model for how two coherence-based agents can deliberate to regulate a domain of interest. We first presented a deductive coherence model, in which the coherence values are derived from the deduction relation of an underlying logic; this allowed us to identify the reasons for why a proposition is accepted or rejected. We then incorporated this coherence model in a model of how agents can generate candidate norms for deliberation, after which we proposed a dialogue protocol for such deliberations. The resulting model was shown to be roughly equally expressive as current logic-based deliberation protocols, while it provides a more natural account of accrual of arguments.

In future research we aim to investigate other possible benefits of coherence models over logic-based argumentation models, as well as formal relations between these models. We also want to study properties of our model, such as the conditions under which an agreement is also internally accepted by the agreeing agents. Finally, we aim to extend the expressiveness of our model, for instance by introducing a distinction between goals and values and by using a richer representation language for norms.

8. REFERENCES

[1] A. Amaya. Inference to the best legal explanation. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*. Ashgate Publishing, Aldershot, 2009. To appear.

[2] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 2009. to appear.

[3] K. Atkinson. *What Should We Do?: Computational Representation of Persuasive Argument in Practical Reasoning*. PhD Thesis, Department of Computer Science, University of Liverpool, Liverpool, UK, 2005.

[4] T. Bench-Capon and H. Prakken. Justifying actions by accruing arguments. In P. Dunne and T. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 247–258, Amsterdam etc, 2006. IOS Press.

[5] T. Bench-Capon and G. Sartor. A quantitative approach to theory coherence. In B. Verheij, A. Lodder, R. Loui, and A. Muntjewerff, editors, *Legal Knowledge and Information Systems. JURIX 2001: The Fourteenth Annual Conference*, pages 53–62, Amsterdam etc, 2001. IOS Press.

[6] J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. van der Torre. The BOID architecture: conflicts between beliefs, obligations, intentions and desires. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*.

[7] A. Casali, L. Godo, and C. Sierra. Graded BDI models for agent architectures. In *Lecture Notes in Computer Science*, volume 3487, 2005.

[8] R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm acceptance. In *The Sixth International Workshop on Agent Theories, Architectures, and Languages*. Springer-Verlag, 1998.

[9] P. Dellunde and L. Godo. *Introducing Grades in Deontic Logics*. LNAI, Springer, to appear., 2008.

[10] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence*, 77:321–357, 1995.

[11] P. Hájek. Metamathematics of fuzzy logic. In *Trends in Logic*, volume 4, 1998.

[12] S. Joseph, C. Sierra, and M. Schorlemmer. A coherence based framework for institutional agents. In *Lecture Notes in Computer Science*, volume 4870, 2007.

[13] S. Joseph, C. Sierra, M. Schorlemmer, and P. Dellunde. Formalising deductive coherence: An application to norm evaluation. In *Normas'08(Extended version), Technical Report(RR-IIIA-2008-02)*, 2009. http://www.iiia.csic.es/ sierra/papers/2009/Coherence.pdf.

[14] P. Pasquier, N. Andrillon, M.-A. Labrie, and B. Chaib-draa. An exploration in using cognitive coherence theory to automate bdi agentsÍ communicational behavior. In *Advances in Agent Communication*. Springer, 2004.

[15] P. Pasquier and B. Chaib-draa. The cognitive coherence approach for agent communication pragmatics. In *Second International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003.

[16] H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, New York, 2005. ACM Press.

[17] A. S. Rao and M. P. Georgeff. Bdi agents: From theory to practice. In *ICMAS–95, First International Conference on Multi-Agent Systems: Proceedings*, pages 312–319. MIT Press, 1995.

[18] P. Thagard. *Coherence in Thought and Action*. MIT Press, 2002.

[19] P. Thagard. Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, 18:231–249, 2004.

[20] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.