Formalizing Deductive Coherence: An Application to Norm Evaluation

Sindhu Joseph[1], Pilar Dellunde[1,2], Marco Schorlemmer[1], and Carles Sierra[1]

[1]Artificial Intelligence Research Institute, IIIA-CSIC, [2]Univ. Autònoma de Barcelona
Bellaterra (Barcelona), Catalonia, Spain
{joseph,pilar,sierra,marco}@iiia.csic.es

Abstract. In this paper we study how agents can autonomously deliberate on norms and cognitions in the context of a normative multiagent system. We propose Thagard's cognitive theory of coherence as a tool to achieve this autonomous deliberation. Taking a proof-theoretic approach, we first provide a formalization of coherence theory, focusing on a particular type of coherence, namely deductive coherence. We then propose a mechanism to compute coherence values between nodes in a coherence graph, making it fully computational. We further introduce a semantic interpretation of coherence using the notion of degrees of consistency by Ruspini. Finally, we illustrate the formalism in a normative multiagent setting where the norms are established to share a common resource, in this case water. We use graded logic to incorporate uncertainty reasoning in our example.

Keywords: Deductive Coherence, Multiagent Systems, Normative Systems

1 Introduction and Motivations

A normative multiagent system is a multiagent system where the agent interactions are governed by norms. In such a system, agents have the explicit right over adherence to the norms [3]. However, there is a lack of mechanisms for autonomous norm evaluations in agents. We try to address this lack for obligatory norms, norms that restrict agents behavior.

Recent works study the relation between agents and norms, in particular the architectural implications of autonomous normative agents [4, 6, 17, 2]. Conte et al. make the process of norm adoption explicit by specifying what it means for agents to adopt a norm, differentiating norm recognition from norm acceptance [6]. However they leave open the various ways an agent can actually reason about norms. BOID [4] is an architecture specially designed for conflict resolution in the context of norms, however the extend of autonomy provided in BOID is limited to agent types. That is, in this architecture, all agents of a certain type, if faced with the same conflict, would come up with the same resolution. Thus, statically assigning priorities between cognitions and norms will not make an agent truly autonomous. Situations are in general the major cause for reevaluation of cognitions and norms, and hence any mechanism intending to bring about agent autonomy ideally should choose those cognitions and norms that best satisfy the constraints imposed by the situation.

Cognitive coherence [14] is such a theory suggesting that humans accept or reject a cognition (external or internal) depending on how much it contributes to maximizing the constraints imposed by situations and other cognitions. Pasquier et al. [10] introduced the possibility of extending agent reasoning with Thagard's theory of coherence. While their contribution introduces the concept of coherence in the field of multiagent systems, a detailed and formal treatment of cognitive coherence is still called for.

According to coherence theory, there are coherence and incoherence relations between concepts depending on whether the concepts support (a positive constraint) or contradict (a negative constraint) each other. If two concepts are not related then there is no coherence (constraint) between them. Normally a graph with nodes[1], and weighted edges are used to represent the set of concepts, and constraints between pairs of concepts. Given such a coherence graph, Thagard defines a mechanism to compute coherence based on maximizing constraint satisfaction, where certain principles are defined to characterize and differentiate different types of coherence relations. Understanding these principles and deducing methods to compute the coherence values between concepts is vital for it to be useful in any application of the theory. Without this important formalization, practical realizations of coherence are hard to imagine.

In this paper we have chosen to analyze one such type of coherence, namely the deductive coherence, because the theorems of logical deduction from which it is derived are well understood. Our aim is to generate coherence values between concepts (in this case, formulas in a language) by formalizing the relationship between coherence and logical entailment. Coherence as a logical relation is significant in itself and has important implications. One of the properties of coherence that makes it different from other branches of logic is its tolerance to inconsistencies. In addition, our setting allows us to work with deductive systems without structural rules such as weakening. We introduce also a semantic characterization of the notion of deductive coherence in terms of similarities between possible worlds [12].

In summary, we advance the state of the art, by introducing coherence as a means to reason about norms in normative multiagent systems. We provide a proof-theoretic account of cognitive coherence, based on Thagard's characterization of deductive coherence [14] (Section 3), and a semantic interpretation of coherence using Ruspini's degrees of consistency [12] (Section 4). Our approach differs from previous formalizations of coherence in the fact that we introduce a fully computational model of coherence, and that we use graded logic to incorporate uncertainty reasoning [5]. Finally our formalization is independent of the underlying logic, that is, given a fixed logic that represents the agent's mental states, the deductive relation of this logic allows us to calculate a degree of coherence between the different sentences of this language. In Section 2 we briefly provide the background on the coherence framework we have proposed in earlier papers [9, 8]. Section 5 is devoted to illustrating how the coherence formalism can be used in a normative multiagent system for autonomous norm evaluation. In Section 6 we compare and contrast our work with that of the state of the art and conclude with a few lines on our future work.

[1] weighted nodes in the case of graded cognitions

2 Coherence Graphs

In this section, we provide the basic definitions of the coherence framework introduced in [9, 8]. In particular we go over the definition of coherence graphs, computation of the coherence values, and the selection of concepts from a coherence graph in a way that maximizes the overall coherence. For a more detailed discussion on the coherence framework and its intuitions refer to [9, 8].

Definition 1. *A* coherence graph *is a graph $G = \langle V, E, \alpha, \zeta \rangle$, where*

1. *V is a finite set of nodes representing concepts*
2. *$E \subseteq V \times V$ is a finite set of edges representing the coherence between concepts.*
3. *$\alpha : V \to [0, 1]$ is a function that maps each node to a weight representing grades (confidence) on the concepts*
4. *$\zeta : E \to [-1, 1] \setminus \{0\}$ is a coherence function that assigns a value to the coherence between concepts.*

According to coherence theory, if a concept is chosen as accepted (or declared true), concepts contradicting it are most likely rejected (or declared false) while concepts supporting it and getting support from it are most likely accepted (or declared true). The important problem is not to find a concept that gets accepted, but to know whether a set of concepts 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 maximize the satisfaction of constraints. A positive constraint between two nodes is said to be satisfied if both nodes are either in the accepted set or both 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 these formally as below:

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

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

In all other cases the constraint is said to be *unsatisfied*.

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

$$S(g, \mathcal{A}) = \frac{\sum_{(v,w) \in C_{\mathcal{A}}} \mid \zeta(v, w) \mid \cdot \alpha(v) \cdot \alpha(w)}{\mid E \mid} \tag{1}$$

Definition 4. *Given a coherence graph $g = \langle V, E, \alpha, \zeta \rangle$ the* coherence of g *is*

$$C(g) = \max_{\mathcal{A} \subseteq V} S(g, \mathcal{A}) \tag{2}$$

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

3 Formalizing Coherence: a Proof-Theoretical Approach

Thagard introduces in [14] the notion of deductive coherence by means of a set of principles:

1. Deductive coherence is a symmetric relation.
2. A proposition coheres with propositions that are deducible from it.
3. Propositions that together are used to deduce some other proposition cohere with each other.
4. The more hypotheses it takes to deduce something, the less the degree of coherence.
5. Contradictory propositions are incoherent with each other.
6. Propositions that are intuitively obvious have a degree of acceptability on their own.
7. The acceptability of a proposition in a system of propositions depends on its coherence with them.

In this section we give a proof-theoretical formalization of the notion of deductive coherence inspired by the principles put forth by Thagard[2]. We base our coherence functions on logical deductive relations , in particular on multiset deductive relations. The concept of a multiset is a generalization of the concept of a set. Intuitively speaking, we can regard a multiset as a set in which the number of times each element occurs is significant, but not the order of the elements. The introduction of multisets in our framework will allow us to deal more adequately with logics as linear logics, relevance logics or multi-valued logics. We denote a"multiset deductive relation" as MDR. We assume that all the MDR we deal with are finitary and decidable. These MDRs are often called *simple consequence relations* [1]. We define both a multiset and MDR in the following.

Definition 5. *A multiset is a pair (A, f) with A a set of formulas of L and $f : A \to \mathbb{N}$ a function from A to the set of positive natural numbers.*

Definition 6. *Given a logical language L, we define a multiset deductive relation (MDR) on a set Σ of formulas of L, as being a binary relation $\vdash$ between finite multisets of formulas of L with the following properties: For all $\Gamma_1, \Gamma_2, \Delta_1, \Delta_2 \subseteq L$ and for all $\gamma \in L$*

*1. **Reflexivity**: $\gamma \vdash \gamma$, for every formula γ*
*2. **Transitivity**: if $\Gamma_1 \vdash \Delta_1, \gamma$ and $\gamma, \Gamma_2 \vdash \Delta_2$, then $\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2$.*

3.1 Coherence Functions

Notation: As usual in sequent calculi, we denote by $\vdash \beta$ the fact that β can be deduced from the empty multiset, and we denote by $\Gamma \vdash$ the fact that the multiset Γ has as consequence the empty multiset. For example, in case that L is classical propositional logic, $\vdash \beta$ means that β is a tautology and $\Gamma \vdash$ means that the multiset Γ is inconsistent.

[2] we do not model principle 3 for now, as in a logic this could mean any proposition is related to any other. However the correct interpretation of this should exempt trivial deductions such as $p, q \vdash p \wedge q$.

We approach the formalization of the deductive coherence by first deriving a coherence function from an MDR. We use Thagard's principles to relate an MDR and the coherence function C. The intuition behind these principles is that whenever two propositions are related by a deductive relation, then there exists a positive coherence between them, the degree of the coherence being inversely proportional to the number of propositions involved in the deduction. If they form a contradiction, then there is a negative coherence between them. We express these in terms of a *support function* SD on the MDR as below.

Definition 7. *Let* $\vdash$ *be a MDR and* $\mathcal{T}$ *a finite set of formulas of the language* L*. A support function* SD *for* $\mathcal{T}$ *is a partial function with*

$$SD(\delta,\beta)=\begin{cases} n+1 & \textit{if there exists } \Gamma\subseteq\mathcal{T} \textit{ with cardinality } |\Gamma|=n \textit{ such that } \Gamma,\delta\vdash\beta \\ & \textit{and } \Gamma,\delta\nvdash \textit{ and } \Gamma\nvdash\beta \textit{ and } |\Gamma| \textit{ is the minimum.} \\ 1 & \textit{if } \vdash\beta \text{ and } \delta\nvdash \\ -1 & \textit{if } \delta,\beta\vdash \end{cases}$$

Observe that, for any given MDR, the support function SD satisfies the following:

1. if $\delta\vdash\beta$, then $SD(\delta,\beta)=1$
2. If $SD(\gamma,\delta)=1$ and $SD(\delta,\beta)=1$, then $SD(\gamma,\beta)=1$
3. In general, if $SD(\gamma,\delta)=n+1$ and $SD(\delta,\beta)=m+1$, then $\max(n,m)+1\leq SD(\gamma,\beta)\leq n+m+1$

Given a MDR $\vdash$ and a finite set $\mathcal{T}$ of formulas of the language L, we define a *deductive coherence function* $\zeta:\mathcal{T}\times\mathcal{T}\rightarrow[-1,1]$ on $\mathcal{T}$ in the following way:

Definition 8. *For any pair* (δ,β) *of formulas in* $\mathcal{T}$*, a* coherence function ζ *is a partial function with*

$$\zeta(\delta,\beta)=\begin{cases} 1/\min(SD(\delta,\beta),SD(\beta,\delta)) & \textit{if both } SD(\delta,\beta) \textit{ and } SD(\beta,\delta) \textit{ are defined} \\ 1/SD(\delta,\beta) & \textit{if } SD(\delta,\beta) \textit{ is defined, } SD(\beta,\delta) \textit{ undefined} \end{cases}$$

For any given MDR, the deductive coherence function ζ is a symmetric function. ζ is not transitive in general, however transitivity holds except in the cases where both $\gamma\vdash\delta$ and $\beta\vdash\delta$ are true or both $\delta\vdash\gamma$ and $\delta\vdash\beta$ are true.

3.2 MDR Coherence Properties

We can classify logics according to structural rules or connectives available in it. There are two types of connectives: the *internal* connectives, which transform a given sequent into an equivalent one that has a special required form, and the *combining* connectives, which combine two sequents into one. For instance, classical propositional logic is monotonic, has the above connectives, and makes no difference between the combining and the corresponding internal connectives. On the other hand, propositional linear logic is nonmonotonic, has the above connectives but distinguish between internal and combining ones. Intuitionistic logic differs from classical propositional logic in its implication connective and does not contain any internal negation. In this section, we study the MDR classifications that give rise to different properties of the coherence functions. We prove certain logical properties of the support and deductive coherence functions over this MDR.

Combining Conjunction: A connective $\wedge$ is a *combining conjunction* iff:

For all $\Gamma, \Delta \subseteq L, \delta, \beta \in L$ we have $\Gamma \vdash \Delta, \delta \wedge \beta$ iff $\Gamma \vdash \Delta, \delta$ and $\Gamma \vdash \Delta, \beta$

By the definition of combining conjunction the following properties hold:

1. If $SD(\gamma, \delta) = 1$ and $SD(\gamma, \beta) = 1$, then $SD(\gamma, \delta \wedge \beta) = 1$
2. If $SD(\gamma, \delta \wedge \beta) = n + 1$, then $0 \leq SD(\gamma, \delta) \leq n + 1$ and $0 \leq SD(\gamma, \beta) \leq n + 1$
3. If $SD(\gamma, \delta \wedge \beta) = 1$, then $SD(\gamma, \delta) = 1$ and $SD(\gamma, \beta) = 1$
4. $SD(\delta \wedge \beta, \delta) = 1$ (whenever $\delta, \beta \nvdash$)

If $\delta, \beta \nvdash$ and $\nvdash \delta$ and $\nvdash \beta$:

1. $\zeta(\delta \wedge \beta, \delta) = 1$
2. If $\zeta(\gamma, \delta \wedge \beta) = 1$, then $\zeta(\gamma, \delta) = 1$ and $\zeta(\gamma, \beta) = 1$ (except when $\delta \wedge \beta \vdash \gamma$)

Internal Conjunction: It is said that a connective $\circ$ is a *internal conjunction* iff:

For all $\Gamma, \Delta \subseteq L, \delta, \beta \in L$ we have $\Gamma, \delta, \beta \vdash \Delta$ iff $\Gamma, \delta \circ \beta \vdash \Delta$

By the definition of internal conjunction the following properties hold:

1. If $SD(\delta \circ \beta, \gamma) = n + 1$, then $0 < SD(\delta, \gamma) \leq n + 2$ and $0 < SD(\beta, \gamma) \leq n + 2$
2. $SD(\delta, \delta \circ \beta) = 2$ (if $\delta, \beta \nvdash$)
3. $\zeta(\delta \circ \beta, \delta) = 1/2$ (if $\delta, \beta \nvdash$)

Combining Disjunction: It is said that a connective $\vee$ is a *combining disjunction* iff:

For all $\Gamma, \Delta \subseteq L, \delta, \beta \in L$ we have $\Gamma, \delta \vee \beta \vdash \Delta$ iff $\Gamma, \delta \vdash \Delta$ and $\Gamma, \beta \vdash \Delta$

By the definition of combining disjunction the following properties hold:

1. If $SD(\gamma, \delta) = n + 1$ or $SD(\gamma, \beta) = m + 1$, then $0 < SD(\gamma, \delta \vee \beta) \leq \min(n + 1, m + 1)$
2. If $SD(\delta, \gamma) = n + 1$ and $SD(\beta, \gamma) = n + 1$, then $SD(\delta \vee \beta, \gamma) \leq n + 1$ (in presence of weakening[3]).
3. If $SD(\delta, \gamma) = 1$ and $SD(\beta, \gamma) = 1$, then $SD(\delta \vee \beta, \gamma) = 1$
4. If $SD(\delta \vee \beta, \gamma) = n + 1$, then $SD(\delta, \gamma) \leq n + 1$ and $SD(\beta, \gamma) \leq n + 1$
5. $SD(\delta, \delta \vee \beta) = 1$

If $\delta, \beta \nvdash$ and $\nvdash \delta$ and $\nvdash \beta$:

1. $\zeta(\delta \vee \beta, \delta) = 1$
2. If $\zeta(\gamma, \delta) = 1$ and $\zeta(\gamma, \beta) = 1$, then $\zeta(\gamma, \delta \vee \beta) = 1$
3. If $\zeta(\delta \vee \beta, \gamma) = 1$, then $\zeta(\delta, \gamma) \leq 1$ and $\zeta(\beta, \gamma) \leq 1$

[3] It is said that a MDR has the *weakening* rule when $\Gamma \vdash \Sigma$ iff $\Gamma, \delta \vdash \Sigma$

Internal Disjunction: It is said that a connective $\circ$ is a *internal disjunction* iff:

For all $\mathit{\Gamma}, \mathit{\Delta} \subseteq L, \delta, \beta \in L$ we have $\mathit{\Gamma} \vdash \mathit{\Delta}, \delta, \beta$ iff $\mathit{\Gamma} \vdash \mathit{\Delta}, \delta + \beta$

By the definition of internal disjunction the following properties hold:

1. If $SD(\gamma, \delta) = n + 1$ or $SD(\gamma, \beta) = m + 1$, then $0 < SD(\gamma, \delta + \beta) \leq \min(n + 1, m + 1)$ (in presence of weakening)
2. If $SD(\delta, \gamma) = n + 1$ and $SD(\beta, \gamma) = m + 1$, then $SD(\delta + \beta, \gamma) \leq n + m + 2$
3. If $SD(\delta, \gamma) = 1$ and $SD(\beta, \gamma) = 1$, then $SD(\delta + \beta, \gamma) = 1$
4. $SD(\delta, \delta + \beta) = 1$ (in presence of weakening)
5. $\zeta(\delta + \beta, \delta) = 1$ (in presence of weakening, if $\delta, \beta \nvdash$)

Combining Implication: It is said that a connective $\supset$ is a *combining implication* iff:

For all $\mathit{\Gamma}, \mathit{\Delta} \subseteq L, \delta, \beta \in L$ we have $\mathit{\Gamma}, \delta \supset \beta \vdash \mathit{\Delta}$ iff $\mathit{\Gamma} \vdash \mathit{\Delta}, \delta$ and $\mathit{\Gamma}, \beta \vdash \mathit{\Delta}$.

By the definition of combining implication the following properties hold:

1. If $SD(\gamma, \beta) = m + 1$, then $0 < SD(\gamma, \delta \supset \beta) \leq m + 1$
2. If $SD(\delta \supset \beta, \gamma) = n + 1$, then $SD(\beta, \gamma) \leq n + 1$
3. $SD(\beta, \delta \supset \beta) = 1$
4. $SD(\delta \supset \beta, \beta) = 2$ (in presence of weakening).

If $\delta, \beta \nvdash$ and $\nvdash \delta$ and $\nvdash \beta$:

1. $\zeta(\delta \supset \beta, \beta) = 1$
2. If $\zeta(\gamma, \beta) = 1$, then $\zeta(\gamma, \delta \supset \beta) = 1$ (except for the case when $\beta \vdash \gamma$)
3. If $\zeta(\gamma, \delta \supset \beta) = 1$, then $\zeta(\gamma, \beta) = 1$ (except for the case when $\gamma \vdash \delta \supset \beta$)

Internal Implication: It is said that a connective $\rightarrow$ is a *internal implication* iff:

For all $\mathit{\Gamma}, \mathit{\Delta} \subseteq L, \delta, \beta \in L$ we have $\mathit{\Gamma}, \delta \vdash \beta$ iff $\mathit{\Gamma} \vdash \delta \rightarrow \beta$

By the definition of internal implication the following properties hold:

1. If $SD(\gamma, \beta) = m+1$, then $0 < SD(\gamma, \delta \rightarrow \beta) \leq m+1$ (in presence of weakening)
2. If $SD(\gamma, \beta) = 1$, then $SD(\gamma, \delta \rightarrow \beta) = 1$ (in presence of weakening)
3. If $SD(\gamma, \delta \rightarrow \beta) = n + 1$, then $SD(\gamma, \beta) \leq n + 2$
4. If $SD(\beta, \gamma) = n + 1$, then $SD(\delta \rightarrow \beta, \gamma) \leq n + 2$
5. $SD(\delta \rightarrow \beta, \beta) = 2$

If $\delta, \beta \nvdash$ and $\nvdash \delta$ and $\nvdash \beta$:

1. $\zeta(\delta \rightarrow \beta, \beta) = 1/2$
2. If $\zeta(\gamma, \delta \rightarrow \beta) = 1$, then $\zeta(\gamma, \beta) = 1/2$ (except when $\gamma, \delta \rightarrow \beta \vdash \gamma$))

Involutive Negation: It is said that a negation is *internal* iff:

For all $\Gamma, \Delta \subseteq L, \delta \in L$ we have $\Gamma, \delta \vdash \Delta$ iff $\Gamma \vdash \Delta, \neg\delta$

Internal negations are *involutive* (that is, for every formula δ, $\delta \vdash \neg\neg\delta$ and $\neg\neg\delta \vdash \delta$). In this case, we have $SD(\delta, \neg\neg\delta) = 1$ and $SD(\neg\neg\delta, \delta) = 1$ and hence $\zeta(\delta, \neg\neg\delta) = 1$. Let us assume that $\delta, \beta \nvdash$ and $\nvdash \delta$ and $\nvdash \beta$. For internal negations we have:

1. $SD(\delta, \beta) = -1$ iff $SD(\delta, \neg\beta) = 1$
2. $SD(\delta, \neg\beta) = -1$ iff $SD(\delta, \beta) = 1$
3. If $\zeta(\delta, \beta) = -1$, then $\zeta(\delta, \neg\beta) = 1$
4. If $\zeta(\delta, \neg\beta) = -1$, then $\zeta(\delta, \beta) = 1$
5. If $\zeta(\delta, \neg\beta) = 1$, then either $\zeta(\delta, \beta) = -1$ or $\zeta(\neg\delta, \neg\beta) = -1$
6. If $\zeta(\delta, \beta) = 1$, then either $\zeta(\delta, \neg\beta) = -1$ or $\zeta(\neg\delta, \beta) = -1$

4 Formalizing Coherence: a Semantical Approach

In this section we propose a semantical formalization of coherence using the notion of *degrees of consistency* introduced by Ruspini in [12]. Ruspini in his work interprets the similarity between two propositions, by the similarity between the worlds in which the propositions are true. Using this interpretation, we define coherence as the similarity between possible worlds.

We first introduce the basic definitions from Ruspini's degree's of consistency and then define coherence in terms of it. For the sake of clarity we restrict now our attention to propositional languages. Let L be a propositional language and W a set of classical interpretations of L (i.e., a set of possible worlds). For any $w \in W$ and any proposition $p \in L$, we denote by $w \models p$ the fact that proposition p is true in the interpretation w. First we introduce some basic definitions.

Definition 9. *A function* $T : [0, 1] \times [0, 1] \to [0, 1]$ *is a* triangular norm *if and only if:*

1. *T is commutative and associative*
2. *T is non-decreasing in both arguments*
3. *$T(1, x) = x$ and $T(0, x) = 0$ for all $x \in [0, 1]$*

Definition 10. *Given a triangular norm T, $S_T : W \times W \to [0, 1]$ is a* T-similarity function *if and only if S_T satisfies the following properties: For all $w, w', w'' \in W$*

1. *Reflexivity: $S_T(w, w) = 1$*
2. *Symmetry: $S_T(w, w') = S_T(w', w)$*
3. *T-Transitivity: $S_T(w, w') \geq T(S_T(w, w''), S_T(w'', w'))$*

where T is a triangular norm function that is continuous (t-norm for short).

The function assigns a degree of similarity between 0 (corresponding to maximum dissimilarity) and 1 (corresponding to maximum similarity). For the sake of simplicity, S_T is required to fulfill that $S_T(w, w') = 1$ implies $w = w'$. The transitivity requirement allows S_T to become a generalized equivalence relation.

Ruspini generalizes the semantical entailment relationship between propositions. He defines both an implication function and a consistency function between propositions. The definition of partial implication between propositions is based on conditions that determine whether, given two propositions p and q, one of them implies the other to the degree n. Observe that the degree of consistency Con is a symmetric measure while the degree of implication Imp is not. Nevertheless, Imp has the T-transitivity property of similarity. Moreover, for any formulas $p, q \in L$, $Con(p \mid q) \geq Imp(p \mid q)$. We introduce the formal definitions below:

Given a T-similarity relation S_T and propositions $p, q \in L$, the *degree of implication* $Imp(p \mid q)$ is defined as:

$$Imp(p \mid q) = \inf_{w' \models q} \sup_{w \models p} S_T(w, w')$$

and Ruspini introduces also the *degree of consistency* $Con(p \mid q)$ in the following way:

$$Con(p \mid q) = \sup_{w' \models q} \sup_{w \models p} S_T(w, w')$$

By definition of the implication and consistency measures it is easy to check that $Imp(p \mid q) = 1$ iff $q \models p$ whereas $Con(p \mid q) = 1$ iff $q \not\models \neg p$. Now we state some basic properties of the consistency degree for L with $p, q, r \in L$ and $n, m \in [0, 1]$:

1. $Con(p \wedge q \mid q) = 1$ iff $p, q \not\vdash$
2. $Con(p \vee q \mid q) = 1$ iff $p, q \not\vdash$
3. If $Con(r \mid p) = n$ and $Con(r \mid q) = m$, then $Con(r \mid p \vee q) = \max(n, m)$
4. If $Con(r \mid p \wedge q) = 1$ then $Con(r \mid p) = 1$ and $Con(r \mid q) = 1$
5. If $Con(r \mid p \vee q) = n$ then $Con(r \mid p) \leq n$ and $Con(r \mid q) \leq n$
6. $Con(p \mid \neg p) = 0$

Now we can define a *coherence function* $C' : \mathcal{T} \times \mathcal{T} \to [-1, 1]$ on $\mathcal{T}$ in terms of degrees of consistency as follows:

Definition 11. *For any pair* (p, q) *of formulas in* $\mathcal{T}$*, a* coherence function $C' : \mathcal{T} \times \mathcal{T} \to [-1, 1]$ *on* $\mathcal{T}$ *is* $C'(p, q) = Con(p, q)$

The relationship between consistency and coherence is a subject of our future work.

5 Example - Norm evaluation

We apply the formalism developed in the previous Sections to model norm evaluation in a real scenario. The example is motivated by the water sharing treaty signed between the southern states of India during 1892 and 1924 and the disputes thereafter [16]. The objectives of this example are threefold. First, to demonstrate how self-interested agents working together evaluate norms. Second, to show the need for *norm adaptation* inspired by individual coherence evaluations, whereas the grander aim is to set up a framework for norm adaptation itself, which will be our future work. And third, to open new application areas in norm evaluation where such cognitive theories could be applied.

We simplify the case for brevity, considering just two agents s and t standing for two distinct Indian states. We model the reasoning of s in two snapshots of time, one when the first treaty is about to be signed (i.e, the decision to adopt the norm) and the second after a period of working together, when the situation has evolved.

5.1 Coherence Maximizing Agent

We describe now the reasoning performed by a coherence maximizing agent. Our agents have graded cognitions [5], as it gives a more realistic representation of agent cognitions, agents often have uncertainty about their cognitions. Hence $B(p, d)$ means agent believes that proposition p is true (in a near future world[4]) with probability d. ($D(p, d)$, and $I(p, d)$ are desires and intentions and are interpreted analogously). Proposition p is a statement about a world and is expressed as triples of the form $\langle object, attribute, value\rangle$. For instance $\langle urbanization, growth\_index, high\rangle$ states that *there is a high growth in urbanization*. The probability degree d of a compound formula is derived from those of its constituents using the formalism as in [5].

Further we have a multi-context (MC) agent [5] which contains three basic components: units or contexts (for the cognitive agents considered here, the contexts are C_b, C_d and C_i corresponding to the belief, desire and intention cognitions), logics, and bridge rules that channel the propagation of consequences between contexts. Hence an MC specification of an agent is a group of interconnected units: $\langle\{C_i\}, \Delta_{br}\rangle$. Each context is a tuple, $C_i = \langle L_i, A_i, \Delta_i\rangle$ where L_i, A_i and Δ_i are the language, axioms, and inference rules respectively. Δ_{br} is the set of bridge rules, which function as inference mechanisms between contexts. In our extension of MC, each context is associated with a coherence graph. We further extend bridge rules to operate on the coherence graphs so that we can perform inferences between graphs and combine graphs. For the details, refer [9]. For example, a $\frac{b:B(\psi,\delta),d:D(\psi,\beta)}{i:I(\psi,\min(\delta,\beta))}$ we introduce in the coherence graph of the cognitions $g = \langle V, E, \alpha, \zeta\rangle$ the following:

- add nodes $I(p_{17}, 0.95)$ if $B(p_{17}, 0.95), D(p_{17}, 0.95)$ to V
- $\alpha(I(p_{17}, 0.95) = 0.95$
- add edges $\{(B(p_{17}, 0.95), I(p_{17}, 0.95))$, and $(D(p_{17}, 0.95), I(p_{17}, 0.95))\}$ to E
- $\zeta(B(p_{17}, 0.95), I(p_{17}, 0.95)) = 0.3, \zeta(D(p_{17}, 0.95), I(p_{17}, 0.95)) = 0.3$

The bridge rules we use in the water-sharing example are $br_1 = \frac{i:I(\psi,\delta)}{d:B(\psi,\delta)}$ and $br_2 = \frac{b:B(\psi,\delta),d:D(\psi,\beta)}{i:I(\psi,\min(\delta,\beta))}$. However, the bridge rules chosen here are only indicative and depend on the agent types that one wants to model.

In our implementation we use a Prolog-based meta interpreter to extract proofs of each sentence in the BDI base of the agent where these proofs will give raise to the coherence values between pairs of sentences using the support function SD of Section 3. We further use a semi-definite programming max-cut approximation algorithm to evaluate the coherence of the graph and to determine the nodes in the accepted set [15].

[4] In our representation we refer to future worlds as the agent is trying to anticipate the coherence of future worlds where the norm is accepted or rejected.

5.2 Norm Adoption

Year : 1892
Agent : s
Action: Evaluating the proposal of the water sharing treaty.
Facts: s is under considerable threat and is not fully autonomous.
Norm to be evaluated: agent s should release 300 billion ft^3 of water to agent t annually.

The agent s reasons by injecting into its internal coherence graph $g_1 = \langle V_1, E_1, \alpha_1, \zeta_1 \rangle$, the anticipated consequences of the norm adoption and compares its coherence on signing the treaty as opposed to not accepting it. Here we use coherence as the primary mechanism for decision making, however in the future we would like to analyze also the influence of sanctions, and rewards. Although in our framework sanctions related to norms are not modeled explicitly, we take into account their influences in forming the agent modalities. Below we list the propositions relevant to forming the agent cognitions and then the cognitions of agent s:

p_{11}	$\langle river\_basin, water\_index, adequate \rangle$
p_{12}	$\langle rain\_fall, index, good \rangle$
p_{13}	$\langle water_release, quantity,$300 billion ft$^3\rangle$
p_{14}	$\langle s_2\_threat, type, military\_force \rangle$
p_{15}	$\langle s_2\_threat, status, realized \rangle$
p_{16}	$\langle norm\_proposal, status, accepted \rangle$
p_{17}	$\langle internal\_demand, status, satisfied \rangle$

- Beliefs: $\{B(p_{11}, 0.90)$,$B(p_{12}, 0.75)$, $B(p_{14}, 1)$, $B(p_{16}, 1)$, $B(p_{11} \wedge p_{12} \wedge p_{13}) \rightarrow p_{17}, 1)$, $B(p_{14} \wedge \neg p_{16} \rightarrow p_{15}, 1)$, $B(p_{16} \rightarrow \neg p_{15}, 1)\}$
- Desires: $\{D(p_{17}, 0.95), D(\neg p_{15}, 1)\}$
- Intentions:$\{I(p_{17}, 0.95), I(\neg p_{15}, 1)\}$

Below we analyze the hypothetical reasoning that agent s does to evaluate the norm, *signing of the treaty* i.e p_{16}.

Case 1: s accepts signing the treaty. Accepting to sign the treaty is equivalent to incorporating an additional belief that at a near future world, p_{16} is true with probablity 1. That is $V_1 := V_1 \cup \{B(p_{16}, 1), I(p_{16}, 1)\}$. Below we calculate the coherence of the agent in conjunction with this additional cognition. Applying the *max-cut* algorithm, we have one of the coherence maximizing partition $(\mathcal{A}, V \setminus \mathcal{A})$ as shown in the Figure 1. The corresponding coherence of the graph, $C(g_1)$ is $4.41/16 = 0.28$.

Case 2: s rejects signing the treaty. The differences if s decides not to accept the norm are that it has the additional belief $B(p_{15}, 1)$ whereas it removes the intention $I(\neg p_{15}, 1)$ as it is reasonable to assume that agent t will realize the threat upon rejecting the treaty. That is $V_1 := V_1 \cup \{B(\neg p_{16}, 1), B(p_{15}, 1)\} \setminus \{I(\neg p_{15}, 1)\}$. With these changes, we have the the coherence of the graph as $C(g_1) = 3.07/16 = .19$. As a coherence agent seeks coherence maximization, s prefers to adopt the norm guided by its coherence value. However we do not rule out the possibilities of other considerations of the agent that can influence its final decision.

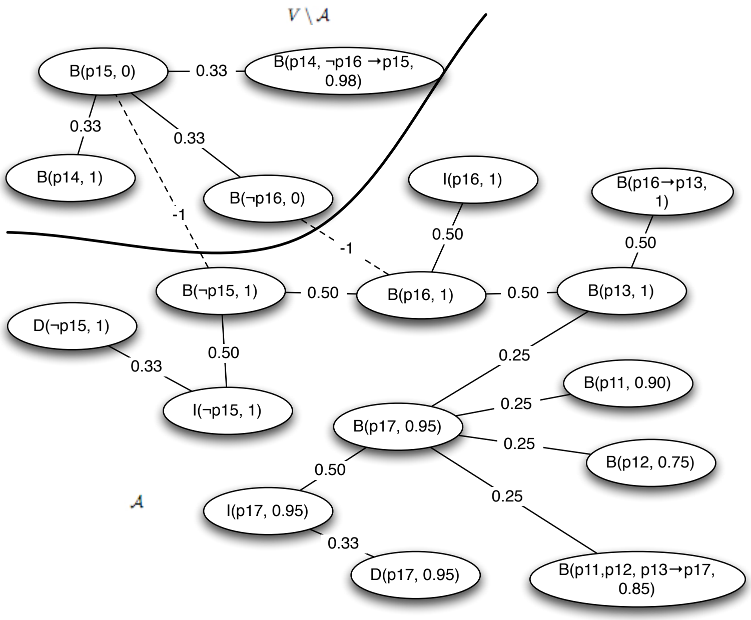

Fig. 1. Coherence graph (g_1), with norm accepted $C(g_1) = 0.28$

5.3 The incoherence buildup

Year : 1991
Agent : s
Action: Updating cognitive graph based on situation change.
Facts: s experiences large-scale industrialization, urbanization, higher water usage, threat from t to obey the norm, and less amount of rain fall.
Below we list the propositions capturing this change in situation and the changed cognitions of the agent s:

p_{21}	$\langle urbanization, growth\_index, high\rangle$
p_{22}	$\langle industrialization, growth\_index, high\rangle$
p_{23}	$\langle water\_usage, growth\_index, high\rangle$
p_{24}	$\langle revenue, growth\_index, high\rangle$

- Beliefs: $\{B(\neg p_{11}, 0.90), B(\neg p_{12}, 0.75), B(p_{14}, 0.75), B(p_{21}, 0.90), B(p_{22}, 0.90),$ $B(p_{23}, 0.95)$, $B(p_{13}, 1)$, $B(p_{16}, 1)$, $B(\neg p_{11} \wedge \neg p_{12} \wedge p_{23} \wedge p_{13} \rightarrow \neg p_{17}, 0.90)$, $B(p_{14} \wedge \neg p_{16} \rightarrow p_{15}, 0.75)$, $B(p_{21} \wedge p_{22} \rightarrow p_{23}, 1)$, $B(p_{24} \rightarrow p_{21}, 1)$, $B(p_{24} \rightarrow p_{22}, 1)$, $B(p_{17} \rightarrow p_{24}, 0.75)$, $B(p_{16} \rightarrow \neg p_{15}, 1)$, $B(p_{24} \rightarrow p_{23}, 0.80)\}$
- Desires: $\{D(p_{17}, 0.95), D(p_{24}, 0.85), D(\neg p_{15}, 1)\}$
- Intentions: $\{I(p_{17}, 0.95), I(p_{24}, 0.85), I(\neg p_{15}, 1), I(p_{16}, 1)\}$

The coherence graph g_2 of the agent s with changed cognitions is shown in Figure 2. Some of the cognitions that do not influence the result have not been included in g_2 for

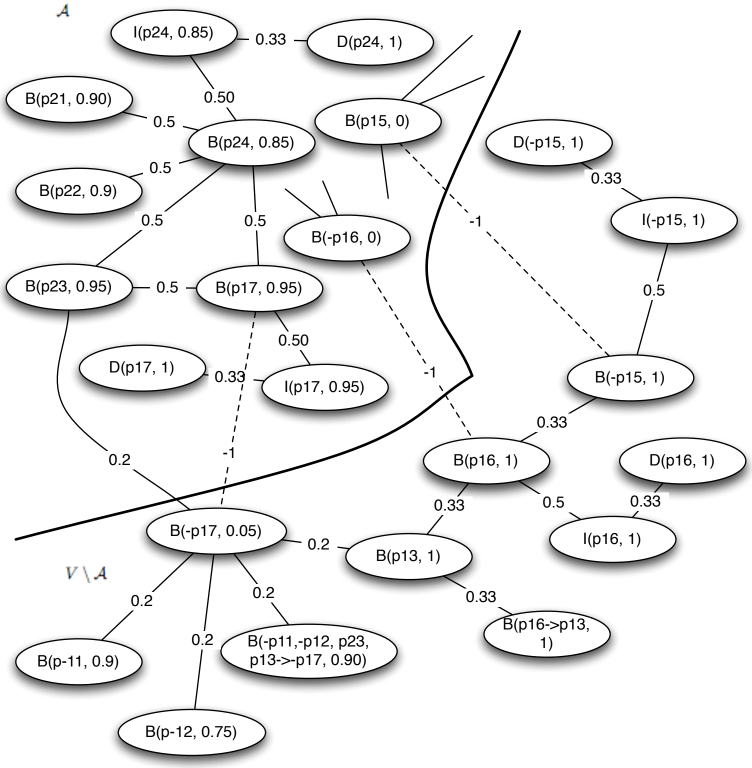

Fig. 2. coherence graph (g_2), $C(g_2) = .29$

the sake of clarity. Using the coherence equations, the coherence maximizing partition $(\mathcal{A}, V \setminus \mathcal{A})$ is shown in Figure 2. The partition interestingly places the cognitions about p_{24} and p_{17} in set $\mathcal{A}$ while the cognitions about p_{16} and $\neg p_{15}$ in set $V \setminus \mathcal{A}$. It is clear from the coherence evaluation that all these intentions cannot coexist while maintaining the maximum coherence. That is the agent has to choose between *obeying the norm, hence avoiding the threat of military action* and *satisfying the internal demands for water, hence economic progress*. Even though the ultimate decision can vary from other considerations of the agent, a purely coherence maximizing agent will choose to violate the norm in order to keep a maximal state of coherence. With this example we show how a coherence maximizing agent evaluates norms in the context of its cognitions.

5.4 Discussion

Even though the example only demonstrates the case of a single norm, the same can be extended to cases where there are multiple norms and there is a need to choose among the norms. In terms of coherence, this is selecting a norm which maximizes the coherence of the graph. By performing the hypothetical analysis of a norm being accepted, norms can be ordered according to the coherence each would generate in the resulting adoption. Another point to note is that here we have assumed our agents to be coherence maximizing. But in reality there are other criteria that need to be considered. Some of them already mentioned and represented in the graph are sanctions and rewards. Another important factor by which an agent makes a decision to adopt a norm is observing the behavior of other agents. We can represent this as cognitive models of other agents.

6 Related and Future work

As discussed in the introduction, the work of Pasquier et al. [10] proposes an agent reasoning theory based on cognitive coherence. The authors have developed a computational model of cognitive coherence based on Thagard's theory of coherence [14]. Thagard in his characterization of coherence, differentiates types of coherence that needs to be accounted for in order to formalize coherence. In our proposal we develop further this idea and take the first step in this direction by giving a proof-theoretic characterization of coherence. Our approach differs from Pasquier et al. as our research is centered on calculating coherence measures, without which a computational model of coherence is hard to realize. Further we understand coherence as a tool not only for maintaining the cognitions of individual agents but also for that of an agent society.

The work of Piwek [11] attempts to model dialogue coherence in terms of generative systems based on natural deduction. The main argument in the paper is that it is possible to generate coherent dialogues by relying on entailments in the agents knowledge base. The paper primarily deals with information seeking dialogue where the definition of whether an agent knows a fact a is equated to whether a can be logically entailed. This is an interesting way to look at dialogue coherence in which the concern is semantic rather than structural. However, here the properties of cognitive coherence as a relation are neither exploited nor modeled. The coherence in this paper refers to the meaning of the term in a linguistic sense, i.e, what makes a text or conversation semantically meaningful, whereas the coherence we deal with is a property of the cognitive state. Though coherence is related to entailment, coherence is not equivalent to it, and it is important to capture and model the differences.

The work of Valencia and Sansonnet [13] models agent dialogue based on the theory of dissonance [7]. This paper exploits the drive to reduce dissonance as a cause to initiate and terminate dialogues. It is curious to note that many authors who have used the theory of dissonance in dialogue initiation and termination [10, 13] have not considered the fact that not all incoherences are dissonances, but dissonance seeks out specialized information or actions. The most important difference between this paper and ours is that for them coherence (or the lack of it) is a local phenomenon concerning only the new arriving fact and the fact that it contradicts with, whereas for us coherence is a global phenomenon affecting the entire knowledge base of the agent. As in the case of Piwek, the authors equate coherence with logical entailment.

As part of our future work, we plan to develop how norm coherence can be analyzed in a normative agent society, study how agents can agree upon or adapt norms. We also would like to explore the semantic interpretation of coherence which is introduced in this paper. Finally we also would investigate more into the example scenario presented in the paper.

Acknowledgements The authors wish to express their thanks to Francesc Esteva, for his comments and ideas on the semantic approach to coherence and the reviewers of this paper for their helpful comments. The research is partially supported by the Open-Knowledge[5] Specific Targeted Research Project (STREP), which is funded by the European Commission under contract number FP6-027253, the Generalitat de Catalunya, under grant 2005-SGR-00093, the Spanish projects "Agreement Technologies" (CONSOLIDER CSD2007-0022, INGENIO 2010) and TIN2006-15662-C02-01. Schorlemmer is supported by a *Ramon y Cajal* research fellowship from Spain's Ministry of Education and Science, which is partially funded by the European Social Fund.

References

1. Arnon Avron. Simple consequence relations. *Inf. Comput.*, 92(1), 1991.
2. Guido Boella and Leendert van der Torre. Fulfilling or violating obligations in normative multiagent systems. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'04)*, 2004.
3. Guido Boella, Leendert van der Torre, and Harko Verhagen. Introduction to normative multiagent systems. In *Normative Multi-agent Systems*, 2007.
4. Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre. The BOID architecture: conflicts between beliefs, obligations, intentions and desires. In *AGENTS '01*, 2001.
5. Ana Casali, Llus Godo, and Carles Sierra. Graded BDI models for agent architectures. In *Lecture Notes in Computer Science*, volume 3487, 2005.
6. Rosaria Conte, Cristiano Castelfranchi, and Frank Dignum. Autonomous norm acceptance. In *ATAL '98*. Springer-Verlag, 1998.
7. Leon Festinger. *A theory of cognitive dissonance*. Stanford University Press, 1957.
8. Sindhu Joseph, Carles Sierra, and Marco Schorlemmer. A coherence based framework for institutional agents. In *Lecture Notes in Computer Science*, volume 4870, 2007.
9. Sindhu Joseph, Carles Sierra, Marco Schorlemmer, and Pilar Dellunde. A multi agent system model of coherence with graded logics. In *Technical Report(RR-IIIA-2008-02)*, 2008.
10. Philippe Pasquier and Brahim Chaib-draa. The cognitive coherence approach for agent communication pragmatics. In *AAMAS '03*, 2003.
11. Piwek. Meaning and dialogue coherence: a proof-theoretic investigation. *Journal of Logic, Language and Information*, 16(4), 2007.
12. Enrique H. Ruspini. On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, 5, 1991.
13. Jean-Paul Sansonnet and Erika Valencia. A model for dialog between semantically heterogeneous informational agents. In *EPIA '03, MAAII*, 2003.
14. Paul Thagard. *Coherence in Thought and Action*. MIT Press, 2002.
15. Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Rev.*, 38, 1996.
16. Wikipedia. Kaveri river water dispute — wikipedia, the free encyclopedia, 2008.
17. Fabiola López y López, Michael Luck, and Mark d'Inverno. Constraining autonomy through norms. In *AAMAS '02*, 2002.

[5] http://www.openk.org