Argumentation and persuasion in the cognitive coherence theory

Philippe Pasquier[1], Iyad Rahwan[2,4], Frank Dignum[3], and Liz Sonenberg[1]

[1] University of Melbourne, Australia
[2] British University of Dubai, UAE
[3] Utrecht University, The Netherlands
[4] (Fellow) University of Edinburgh, UK

Abstract. This paper presents a coherentist approach to argumentation that extends previous proposals on cognitive coherence based agent communication pragmatics (inspired from social psychology) and propose (1) an alternative view on argumentation that is (2) part of a more general model of communication. In this approach, the cognitive aspects associated to both the production, the evaluation and the integration of arguments are driven by calculus on a formal characterization of cognitive coherence.

1 Introduction

"Argumentation is a verbal, social and rational activity aimed at convincing [...] of the acceptability of a standpoint by putting forward a constellation of proposition justifying or refuting the proposition expressed in the standpoint." [32, page 1].

In AI and MAS, argumentation frameworks have been put forward for modelling inference, non-monotonic reasoning, decision making and argumentation-based communication has been introduced has a way to refine multiagent communication [23, 15, 7, 6]. The syntax and semantics of argumentation have been extensively studied, but the pragmatics of argumentation (theory of its use in context) has not been inquired. While the conventional aspects of pragmatics have been taken into account in the formalisms proposed for argumentation dialogues, the cognitive aspects of argumentation have been less studied: when does an agent argue, with whom, on what topic? What are the cognitive effects of arguments (in terms of persuasion and integration)? What is the utility of the argumentation? Are the agents satisfied with their dialogue?

Cognitive coherence theory [20, 21, 16] has been put forward as a way to model the cognitive aspects of agent communication pragmatics (section 2). Inspired from social psychology theories, cognitive coherence provides a native yet realistic modelling of the cognitive aspects of communication through the concept of *attitude change* which captures the persuasive aspect inherent to all communications (section 3). In this paper, we extend the cognitive coherence approach to argumentation and show how this extension allows to model

the generative aspect of argumentation communication as well as the cognitive response to persuasive arguments using a single set of principles (section 4). Finally, a discussion relates this new approach to previous proposals (section 5).

While at the beginning of this ongoing research work, this paper extends the state of the art by (1) proposing an alternative (coherentist) view on argumentation that is (2) part of a more general model of communication (including the cognitive aspect of pragmatics) and (3) giving a fully computational characterization of this new model.

2 The cognitive coherence framework

In cognitive sciences, cognitions gather together all cognitive elements: perceptions, propositional attitudes such as beliefs, desires and intentions, feelings and emotional constituents as well as social commitments.

In cognitive or social psychology, most cognitive theories appeal to the concept of homeostasis, i.e. the human faculty to maintain or restore some physiological or psychological constants despite the outside environment variations. All these theories share as a premise the *coherence principle* which puts coherence as the main organizing mechanism: *the individual is more satisfied with coherence than with incoherence*. The individual forms an opened system whose purpose is to maintain coherence as much as possible.

The core of our theoretical model is the unification of the dissonance theory from Festinger [11] and the coherence theory from Thagard [29]. In that context, our main and original theoretical contribution has been to extend that model to communication (which as not been treated by those two theorists) and to develop a formalism suited to MAS.

2.1 Formal characterization of cognitive coherence

While several formal characterizations of cognitive coherence has been made (logic-based [24], neural network or activation network based [26], probabilistic network [30], decision-theoretic, ...), we present one that is constraint satisfaction based resulting in a simple symbolic-connexionist hybrid formalism (we refer the reader to [28] for an introduction to this family of formalisms).

In this approach, cognitions are represented through the notion of elements. We denote $\mathbb{E}$ the set of all elements. *Elements* (i.e. cognitions) are divided in two sets: the set $\mathcal{A}$ of *accepted elements* and the set $\mathcal{R}$ of *rejected elements*. A closed world assumption which states that *every non-explicitly accepted element is rejected* holds. Since all the cognitions are not equally modifiable, a *resistance to change* is associated to each element of cognition. In line with Festinger [11], a cognition's resistance to change depends on its type, age, as well as the way in which it was acquired: perception, reasoning or communication. Resistances to change allow to differentiate between beliefs that came from perception, beliefs that cames from reasoning and beliefs that came from communication as well as to represent the individual commitment strategies associated with individual

intention. Resistance to change can be accessed through the function $Res : \mathbb{E} \longrightarrow \mathbb{R}$.

Those elements can be cognitively related or unrelated. For elements that are directly related, two types of non-ordered binary constraints represent the relations that hold between them in the agent's cognitive model:

- *Positive constraints*: positive constraints represent positive relations like facilitation, entailment or explanatory relations.
- *Negative constraints*: negative constraints stand for negative relations like mutual exclusion and incompatibility relations.

We note $\mathcal{C}^+$ (resp. $\mathcal{C}^-$) the set of positive (resp. negative) constraints and $\mathbb{C} = \mathcal{C}^+ \cup \mathcal{C}^-$ the set of all constraints. For each of these constraints, a weight reflecting the importance degree for the underlying relation can be attributed[5]. Those weights can be accessed through the function $Weight : \mathbb{C} \longrightarrow \mathbb{R}$. Constraints can be satisfied or not.

Definition 1. (Cognitive Constraint Satisfaction) *A positive constraint is satisfied if and only if the two elements that it binds are both accepted or both rejected, noted* $Sat^+(x,y) \equiv (x,y) \in \mathcal{C}^+ \wedge [(x \in \mathcal{A} \wedge y \in \mathcal{A}) \vee (x \in \mathcal{R} \wedge y \in \mathcal{R})]$. *On the contrary, a negative constraint is satisfied if and only if one of the two elements that it binds is accepted and the other one rejected, noted* $Sat^-(x,y) \equiv (x,y) \in \mathcal{C}^- \wedge [(x \in \mathcal{A} \wedge y \in \mathcal{R}) \vee (x \in \mathcal{R} \wedge y \in \mathcal{A})]$. *Satisfied constraints within a set of elements* $\mathcal{E}$ *are accessed through the function* $Sat : \mathcal{E} \subseteq \mathbb{E} \longrightarrow \{(x,y) | x,y \in \mathcal{E} \wedge (Sat^+(x,y) \vee Sat^-(x,y))\}$

In that context, two elements are said to be *coherent* if they are connected by a relation to which a satisfied constraint corresponds. And conversely, two elements are said to be *incoherent* if and only if they are connected by a non-satisfied constraint. These relations map exactly those of dissonance and consonance in Festinger's psychological theory. The main interest of this type of modelling is to allow defining a metric of cognitive coherence that permit to reify the coherence principle in a computational calculus.

Given a partition of elements among $\mathcal{A}$ and $\mathcal{R}$, one can measure the *coherence degree* of a non-empty set of elements $\mathcal{E}$. We note $Con()$ the function that gives the constraints associated with a set of elements $\mathcal{E}$. $Con : \mathcal{E} \subseteq \mathbb{E} \longrightarrow \{(x,y) \mid x,y \in \mathcal{E}, (x,y) \in \mathbb{C}\}$.

Definition 2. (Cognitive Coherence Degree) *The* ***coherence degree*** $C(\mathcal{E})$, *of a non-empty set of elements,* $\mathcal{E}$ *is obtained by adding the weights of constraints linking elements of* $\mathcal{E}$ *which are satisfied divided by the total weight of concerned constraints. Formally:*

$$C(\mathcal{E}) = \frac{\sum_{(x,y) \in Sat(\mathcal{E})} Weight(x,y)}{\sum_{(x,y) \in Con(\mathcal{E})} Weight(x,y)} \quad (1)$$

[5] This is a way of prioritizing some cognitive constraints as it is done in the BOID architecture [4].

The general coherence problem is then:

Definition 3. (Cognitive Coherence Problem) *The general* ***coherence problem*** *is to find a partition of the set of elements into the set of accepted elements $\mathcal{A}$ and the set of rejected elements $\mathcal{R}$ that maximize the cognitive coherence degree of the considered set of elements.*

It is a constraint optimization problem shown to be NP-complete in [31]. An agent can be partially defined as follows:

Definition 4. (Agent) *An* ***agent*** *is characterized by a tuple $\{\mathcal{P}, \mathcal{B}, \mathcal{I}, SC, \mathcal{C}^+, \mathcal{C}^-, \mathcal{A}, \mathcal{R}\}$, where:*

- *$\mathcal{P}$,$\mathcal{B}$,$\mathcal{I}$ are sets of elements that stand for perceptions, beliefs and individual intentions respectively, SC is a set of elements that stand for the agent's agenda, that stores all the social commitments from which the agent is either the debtor or the creditor;*
- *$\mathcal{C}^+$ (resp. $\mathcal{C}^-$) is a set of non-ordered positive (resp. negative) binary constraints over $\mathcal{P} \cup \mathcal{B} \cup \mathcal{I} \cup SC$ such that $\forall (x, y) \in \mathcal{C}^+ \cup \mathcal{C}^-$, $x \neq y$;*
- *$\mathcal{A}$ is the set of accepted elements and $\mathcal{R}$ the set of rejected elements and $\mathcal{A} \cap \mathcal{R} = \varnothing$ and $\mathcal{A} \cup \mathcal{R} = \mathcal{P} \cup \mathcal{B} \cup \mathcal{I} \cup SC$.*

Beliefs coming from perception ($\mathcal{P}$) or from reasoning ($\mathcal{B}$) as well as intentions ($\mathcal{I}$) constitute the *private cognitions* of the agent, while public or social cognitive elements are captured through the notion of social commitments (as defined in [22]). Social commitment has proven to be a powerful concept to capture the interdependencies between agents [27]. In particular, it allows to represent the semantics of agents communications while respecting the principle of the asymmetry of information that indicates that in the general case what an agent say does not tell anything about what he think (but still socially commits him).

This agent model differs from classical agent modelling in that motivational attributes are not statically defined but will emerge from the cognitive coherence calculus. Concretely, this means that we don't have to specify the agent desires (the coherence principle allows to compute them) but only potential intentions or goals. Examples to be given in this paper will highlight the *motivational drive* associated with cognitive coherence.

Incoherence being conceptually close to the notion of conflict, we use a typology borrowed from works on conflicts [8].

Definition 5. (Internal vs. External Incoherences) *An incoherence is said to be* ***internal*** *iff all the elements involved belongs to the private cognitions of the agent, else it is said to be* ***external****.*

2.2 Local search algorithm

Decision theories as well as micro-economical theories define utility as a property of some valuation functions. A function is a *utility function* if and only if it

reflects the agent preferences. In the cognitive coherence theory, according to the afore-mentioned coherence principle, coherence is preferred to incoherence.

In order to try to maximize its coherence, at each step of his reasoning, an agent will search for a cognition acceptance state change which maximizes the coherence increase, taking into account the resistance to change of that cognition (technically a 1-optimal move). If this attitude is a commitment, the agent will attempt to change it through dialogue and if it is a private cognition (perceptions, beliefs or intentions), it will be changed through attitude change.

In our model, an agent determines which is the most useful cognition acceptance state change by exploring all states reachable from its current state and selects the cognition which can *in case of a successful change* be the most useful to change. A state is said to be reachable if it can be obtained from the current state by modifying only one cognition. A notion of cost has been introduced to advocate for the fact that all cognitions cannot be equally modified. All explored states are so evaluated through an *expected utility function*[6], G, expressed as below:

$$\begin{aligned} G(ExploredState) =& C(ExploredState) - C(CurrentState) \\ & - Res(cognitionChanged) \end{aligned}$$

where $ExploredState$ is the evaluated state, $cognitionChanged$ is the cognition we are examining the change, and Res is the function returning the resistance to change of the manipulated element which expresses the cost of the change.

A recursive version of the local search algorithm is presented in Figure 1 and consists of four phases:

1. For each element e in the agent state, calculate the expected utility and the gain (or loss) in coherence that would result from flipping e, i.e. moving it from $\mathcal{A}$ to $\mathcal{R}$ if it is in $\mathcal{A}$, or moving it from $\mathcal{R}$ to $\mathcal{A}$ otherwise.
2. Produce a new solution by flipping the element that most increases coherence, or with the biggest positive expected utility if coherence cannot be improved.
3. Repeat 1 and 2 until either a social commitment is encountered (a dialogue is needed as an attempt to flip it) or until there is no flip that increases coherence and no flip with positive expected utility.
4. Return result. The solution will be applied if and only if the cumulated expected utility is positive.

The local search algorithm is an informed breath first search algorithm with the afore-mentioned expected utility measure as its heuristics. Update of the resistance to change of the modified elements avoid looping. Since our algorithm does not make any backtracking, the complexity of this algorithm is polynomial: $\mathcal{O}(mn^2)$, where n is the number of elements considered and m the number of

[6] Note that our expected utility function does not include any probabilities. This reflect the case of equiprobability in which the agent have no information about others behavior. Notice that integrating algorithms to progressively learn such probabilities is an obvious perspective of the presented model.

constraints that binds them[7]. We don't have a proof of correctness of this algorithm in regards to the general coherence problem but, as [31] (who used it in another context), it behaved well on tested examples. We refer the interested reader to [16] for full justification and discussion of this algorithm. Traces of execution will be provided along with the examples in this paper.

Function LocalSearch(W)

Inputs: $W = \{\mathcal{P}, \mathcal{B}, \mathcal{I}, SC, \mathcal{C}^+, \mathcal{C}^-, \mathcal{A}, \mathcal{R}\}$; // current agent state
Outputs: List, $Change$; // ordered list of elements (change(s) to attempt).
Global:
Local:
Float, G, $Gval$, C, $Cval$; // Expected utility value of the best move;
Elements set, A', R';
Elements, y, x;
Agent, J; // Agent state buffer
Body:
for all $x \in \mathcal{P} \cup \mathcal{B} \cup \mathcal{I} \cup SC$ **do**
  if $x \in \mathcal{A}$ **then**
    $\mathcal{A}' := \mathcal{A} - \{x\}$; $\mathcal{R}' := \mathcal{R} \cup \{x\}$;
  else
    $\mathcal{R}' := \mathcal{R} - \{x\}$; $\mathcal{A}' := \mathcal{A} \cup \{x\}$;
  end if
  $W' := \{\mathcal{P}, \mathcal{B}, \mathcal{I}, SC, \mathcal{C}^+, \mathcal{C}^-, \mathcal{A}', \mathcal{R}'\}$;
  $G := C(W') - C(W) - Res(x)$; // Expected utility of flipping x
  $C := C(W') - C(W)$; // Pure coherence gain
  if $G > Gval$ **then**
    $J := W'$; $y := x$; $Gval := G$; $Cval := C$;
  end if
end for// Ends when (coherence is not raising anymore and the expected utility is not positive) or a social commitment need to be changed.
if ($Cval < 0$ and $Gval < 0$) or $y \in SC$ **then**
  Return $Change$;
else
  Update ($Res(y)$); Add (J,$Change$);
  LocalSearch(J);
end if

Fig. 1. Recursive specification of the local search algorithm.

2.3 Cognitive coherence applied to agent communication

Applied to agent communication, the cognitive coherence theory supplies theoretical and practical elements for automating agent communication. The cog-

[7] n coherence calculus (sum over m constraints) for each level and a maximum of n levels to be searched.

nitive coherence framework provides the necessary mechanisms to answer (even partially) the following questions which are usually poorly treated in the AI and MAS literature:

1. *Why should agents converse?* Agents dialogue in order to reduce incoherences they cannot reduce alone. We distinguish internal (or personal) incoherence from external (or collective) incoherence depending on whose elements are involved in the incoherence[8].
2. *When should an agent take a dialogue initiative, on which subject and with whom?* An agent engages in a dialogue when an incoherence appears or when an incoherence magnitude exceeds a fixed level[9] and he cannot reduce it alone. Whether because it is an external incoherence and he cannot accept or reject external cognitions on his own, or because it is an internal incoherence he fails to reduce alone. The subject of this dialogue should thus focus on the elements which constitute the incoherence. The dialogue partners are the other agents involved in the incoherence if it is an external one or an agent he thinks could help him in the case of a merely internal incoherence.
3. *By which type of dialogue?* Even if we gave a general mapping of incoherence types toward dialogue types using Walton and Krabble typology [20], the theory is generic enough to be applied to any conventional communicational framework. In [21], we gave the procedural scheme for this choice using DIAGAL [19] dialogue games as primitive dialogue types.
4. *How to define and measure the utility of a conversation?* As we state in [17], following the coherence principle and the classical definition of utility functions, the utility of a dialogue is the difference between the incoherence before and after this dialogue minus the cost of the dialogue moves (expressed in term of the resistance to change of the modified elements). Furthermore, we define the expected utility of a dialogue as the incoherence reduction in the case of success of the dialogue, i.e. the expected dialogue results are reached. As dialogues are attempts to reduce incoherence, expected utility is used to choose between different competing dialogues moves (including dialogue initiative and dialogue ending).
5. *When to stop dialogue or, how to pursue it?* The dialogue stops when the incoherence is reduced[10] or, either it continues with a structuration according to the incoherence reductions chain.
6. *What are the impacts of the dialogue on agents' private cognitions?* In cases where dialogue, considered as an attempt to reduce an incoherence by working on the external world, definitively fails, the agent reduces the incoherence by changing his own mental attitudes in order to recover coherence (this is the attitude change process described in section 3).

[8] In the presented system, external elements are social commitments.

[9] This level or a "Should I dialogue ?" function allows us to model different strategies of dialogue initiative.

[10] Note that this ending criterium is to be tempered with other external factors like like time, resources and social norms. Those resources can be taken into account in the update of the resistance to change of various discussed elements.

7. *Which intensity to give to illocutionary forces of dialogue acts?* Evidently, the intensities of the illocutionary forces of dialogue/speech acts generated are influenced[11] by the incoherence magnitude. The more important the incoherence magnitude is, the more intense the illocutionary forces are.
8. *What are the impacts of the dialogue on agents' mood?* The general scheme is that: following the coherence principle, coherence is a source of satisfaction and incoherence is a source of dissatisfaction. We deduce emotional attitudes from internal coherence dynamic (happiness arises from successful reduction, sadness from failed attempt of reduction, fear from a future important reduction attempt, stress and anxiety from an incoherence persistence,...).
9. *What are the consequences of the dialogue on social relations between agents?* Since agents can compute and store dialogue utility, they can build and modify their relations with other agents in regard to their past dialogues. For example, they can strengthen relations with agents with whom past dialogues were efficient and useful, according to their utility measures, ...

All those dimensions of our theory - except 7, 8 and 9 - have been implemented and exemplified as presented and discussed in [17] and [21]. The presented practical framework relies on our dialogue games based agent communication language (DIAGAL) [19] and our dialogue game simulator toolbox (DGS)[5].

3 Attitude change and persuasion.

From the set of all private cognitions result *attitudes* which are positive or negative psychological dispositions towards a concrete or abstract object or behavior.

For contemporary psychologists, attitudes are the main components of cognition. These are the subjective preliminary to rational action [10]. Theoretically, an agent's behavior is determined by his attitudes. The basic scheme highlighted by those research is that beliefs (cognition) and desires (affect) lead to intentions which could lead to actual behaviors or dialogical attempts to get the corresponding social commitments depending on their nature.

From another point of view, it could happen (due to hierarchies, power relations, value-based negotiation, argumentation,...) that an agent comes to accept a counter-attitudinal course of action or proposition. In that case, *attitude change* might occur. Since cognitive coherence theory is built over five decades of research on attitude change in social psychology, it provides a native yet realistic modelling of the cognitive aspects of persuasion through this concept of attitude change. Within our characterization of cognitive coherence, attitude change refers to the change of acceptance states of some private element of cognition in order to restore coherence with external interdependencies, i.e. social commitments.

[11] Actually, this is not the only factor, other factors could also matter: social role, hierarchical positions,...

4 Argumentation in the cognitive coherence theory

Argumentation has not been introduced in the cognitive coherence approach yet. However, this extension follows naturally from previous work by saying that argumentation, explanation and justification are the processes by which an agent shows to the other agents why his (or a given) position is coherent. In that context, we do not distinguish between argumentation, explanation and justification which all aim to convince in some way. More specifically, the idea behind argumentation is that agents can construct, exchange and weigh up arguments relevant to conflicting issues, in the context of an explicit external incoherence.

The argumentation process can be modelled using three steps: (1) argument generation, (2) argument evaluation and (3) argument integration. The next sections present and exemplify how cognitive processes associated with those steps are computed in the cognitive coherence framework.

4.1 Argument generation

Argumentation is a type of information disclosure. While in cooperative systems this information might be useful to help solving conflicts, or by making the negotiation and the convergence to a deal more efficient, it has been shown in [14] that argumentation and full cooperation is not necessarily always the best strategy for negotiation convergence. More generally, it is unclear if such information disclosure is worth in open system were heterogeneous and competitive (even malicious) agents can use this information to endorse non-cooperative behavior. In this paper, we won't address strategic issues related to argumentation.

In our framework, argumentation can be achieved by constraint propagation by introducing a syntactic facility that will allow the agents to send to one another parts of their elements and constraints networks. Previous work has been done around that idea in the field of distributed constraint satisfaction [13, 14].

Definition 6. ***(Argument)*** *An* argument *for an element acceptance or rejection is a set of elements (along with their acceptance states and resistances to change) and constraints (along with their weights) that form a connected component in the network of cognitions of the agent. More formally, an argument w is a pair $w = \langle H, h \rangle$ such that:*

1. $H \subseteq \mathbb{E}, h \in \mathbb{E}; H \cap \{h\} = \varnothing$;
2. $\forall x, y \in H \cup \{h\}, \exists z_1, ..., z_n \in H \cup \{h\}, (x, z_1), ..., (z_n, y) \subseteq \mathbb{C}$ *(connexity condition);*

H is called the support of the argument while h is the conclusion of the argument.

Definition 7. ***(Argument types)***

Arg_X stands for the set of all possible arguments that can be generated from the agent's bases included in X. It is useful to differentiate between:

- *belief arguments: $\langle H, h \rangle$ is a belief argument iff $(H \cup \{h\}) \subset Arg_{\mathcal{P} \cup \mathcal{B}}$;*

- *practical arguments:* $\langle H, h\rangle$ *is a practical argument iff* $(H \cup \{h\}) \subset Arg_{\mathcal{P}\cup\mathcal{B}} \wedge h \in \mathcal{I}$;
- *social arguments:* $\langle H, h\rangle$ *is a social argument iff* $(H \cup \{h\}) \subset Arg_{\mathcal{I}\cup SC} \wedge (H \cup \{h\}) \cap SC \neq \varnothing$;

In the cognitive coherence framework, argumentation will be used when an explicit external incoherence is not solved otherwise (for example by referring to an authority relation or a social norm). When this precondition will be met, the agents will disclose the private part of the connected component related to the discussed issue. Let's take an example to illustrate this argument generation systematics and illustrate previous definitions.

Two agents W and J are driving a car (it is a joint activity and the agents have complementary access to the necessary resources). The car is at a stop and the agents have to decide which way to go. Suppose that the initial states of agents W and J are the ones presented by Figure 2. Since W wants to go left (he has the corresponding intention accepted), he wants the corresponding social commitment to be accepted (see Figure 3). W will thus make an offer to J[12]:

W: I would turn left.

If agent J also would have wanted to turn left (W's proposal would have been coherent with her views), she would have then accepted the proposal and the corresponding social commitment would have been accepted:

J: Ok.

However, as depicted by Figure 2 agent J want to turn right (i.e. the corresponding intention is accepted), W's proposal acceptance would entail a lost in coherence for J (see Figure 3). J will then embed a counter-proposal[13] as attempt to get a result that would be more coherent with her view. Her argument for this choice (j) will be attached to her proposal:

J: There 's a lot of lights on the left road, that will slow us down. Can't we turn right instead?

Notice that, this make the external incoherence explicit for W[14]. In order to complete the argumentation dialogue initiated by J, W will disclose his own argument (w).

[12] More precisely, he will propose to enter an offer game (see [19] for details about the DIAGAL agent language) which is the only game which entry and success conditions unify with the current and wanted state respectively. Using the current framework and algorithms this will result automatically from the situation described by Figure 2 as described in [16]. This is what the cognitive coherence framework is made for: automatizing agent communications.

[13] In the form of a DIAGAL request game.

[14] See [21] and [16] for a discussion about the importance of the explicitation phase of dialogue that is usually neglected.

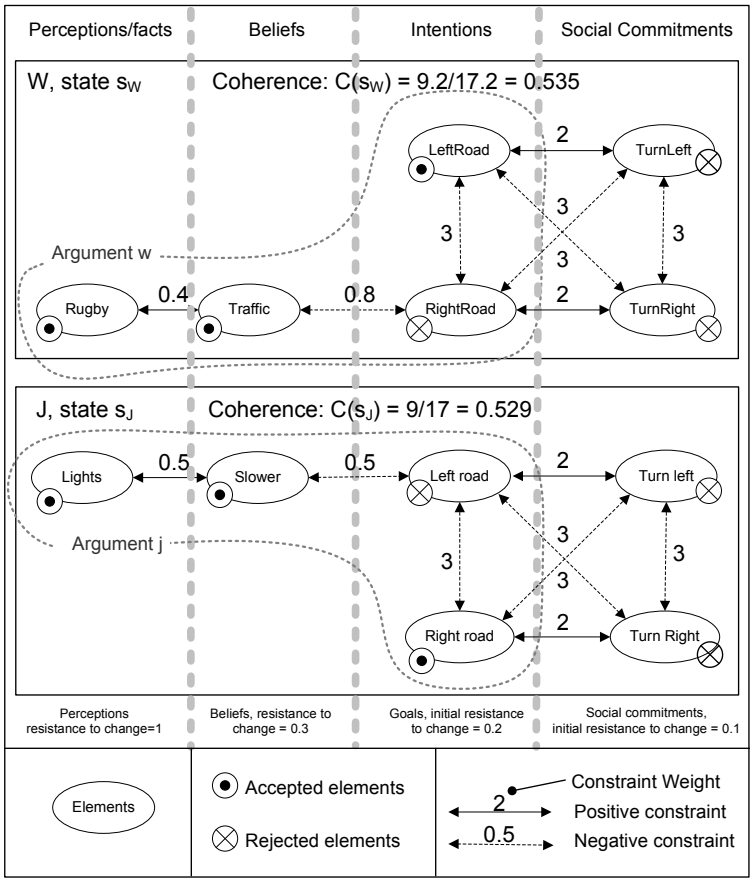

Fig. 2. Initial states s_W and s_J for W and J. Here all the resistances to change are initialized like shown in order to indicate that perceptions are more resistant then beliefs, that are more resistant than intentions that are more resistant than social commitments. Other choices may be made.

W: Yes, but there is a rugby match today, so there will be a lot of traffic on the right road, we should avoid going this way and turn left.

During that process, the agents eventually communicate each other the entire connected component attached to the discussed issues. However, this doesn't tell anything about the way they will evaluate and integrate the exchanged arguments. Next section discusses and proposes a modelling of those dimensions.

4.2 Issues in argument evaluation and integration

Argument evaluation and integration are complex issues, and social psychology (which has studied that problem on experimental basis for half a century now)

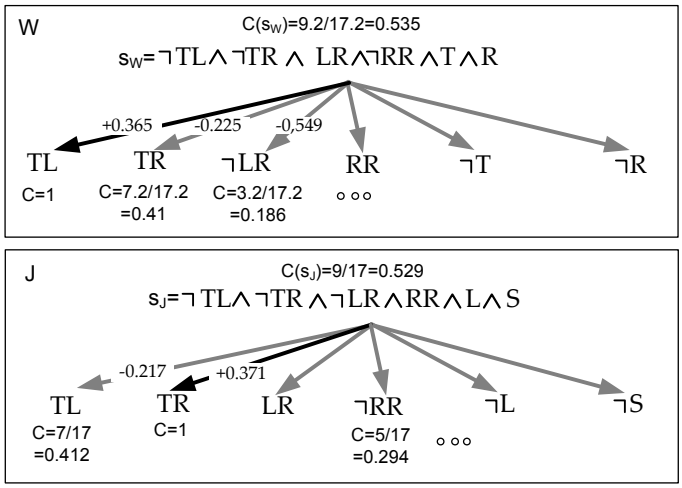

Fig. 3. Reasoning as computed by the local search algorithm from the initial states s_W and s_J for W and J. Here the perceptions/beliefs that "there is a rugby match", "there is a lot of traffic", "there are a lot of lights", "traffic is slower" are noted R, T, L, S respectively, the intentions to turn left and to turn right are noted LR and RR respectfully and the social commitments to turn left and right are noted TR and TL. Rejected elements are noted with a negation sign and only the root of the search tree indicates the full state of the agent, the others nodes just indicate the change they imply. Arcs are labelled with the value of the expected utility function (presented section 2.2). The black path indicates the change(s) returned by the local search algorithm.

indicates that there is a large number of aspects to be considered [10]. Here is a simplified listing of those:

- *evaluation of the source*: authority, trust, credibility, attractiveness;
- *evaluation of the message*: comprehension and quality of argument, number and order of arguments, one- and two-sided messages, confidence, fear;
- *characteristics of the audience*: intelligence and self-esteem, psychological reactance, initial attitudes, heterogeneity, sex differences;
- *characteristics of the medium*: media and channel of communication, media functions, temporality of the communication.

Furthermore, many studies indicates that the regularities in that area are difficult to find and that argumentation evaluation and integration are also linked to cognitive learning and thus depend on the dynamics of the learner [12]. However, a characterization of rational agent argumentation may not take all of these into consideration. We thus restrict the discussion to the salient elements that are already considered in cognitive agent modelling and MAS:

- *trust and credibility*: the levels of trust and credibility associated with the protagonist influence the argument evaluation and integration process. The

model presented in [24] (inspired by cognitive coherence approach) has inquired this link further. For the sake of simplicity, in this paper, we will consider that the level of trust and credibility are the highest possible;
- *initial attitude toward the standpoint defended by the argument*: it is clear that the initial attitude of the antagonist agent will intervene in argument evaluation and integration especially in conjunction with trust and credibility. Social psychology, in particular the theory of social judgment [25], showed that each agent maintain some acceptability intervals in which arguments may be taken into account while arguments falling out of those intervals will be considered too extreme and won't be taken into account. However, because we model rational agents that usually operate in quite precise and well known domain, we will make the assumption that all arguments will be considered;
- *initial attitude toward the protagonist of the argument*: this issue is related to the level of trust and cooperativeness that the antagonist shows toward the protagonist. Will the agent integrate the other's point of view in their own and act accordingly (which would be very cooperative) or will they compare their point of view with the other's and then substitute those two if their is weaker or reject the other's one if it is (subjectively) evaluated as weaker? In this paper, we make the assumption that the agents will fully integrate the other argument in their mental states;
- *Heterogeneity of the participants*: we call *objective evaluation*) the case where all the participants share the same evaluation function and we name *subjective evaluation* the case in which they all have their own. This aspect depend on the type of system addressed. While objective evaluation might be possible in cooperative systems, open system where agents may be heterogeneous will most probably rest on subjective evaluation. In this paper, we will make the assumption that the agents share the same evaluation function to be described.
- *number and quality of arguments*: in this paper, we will focus on cognitive factors which will tend to reduce argument evaluation to this last category.

4.3 Argument evaluation

Argument evaluation will be done by comparing (using a shared measure) the strengths of the arguments provided by both sides in order to decide whose standpoint will be chosen as the more rational one. We use the following argument evaluation measure:

Definition 8. (Strength of an argument)
The strength of a given argument $\langle H, h\rangle$ is the sum of the weights of the satisfied constraints minus the sum of the weights of the non-satisfied ones. Formally:

$$Strengh(\langle H, h\rangle) = 2 * \sum_{(x,y)\in Sat(H\cup h)} Weight(x, y) - \sum_{(x,y)\in Con(H\cup h)} Weight(x, y)$$

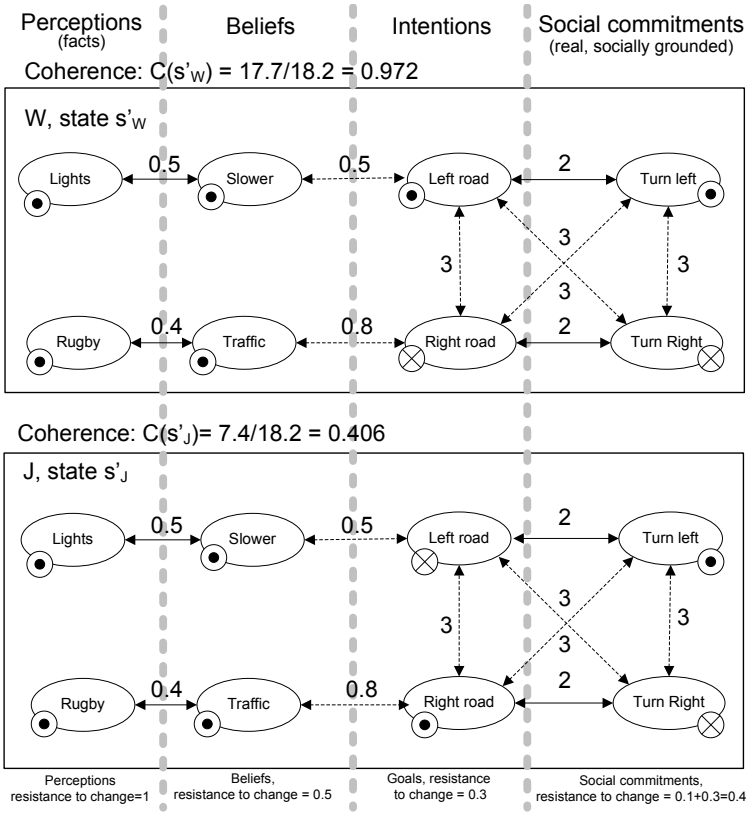

Fig. 4. W and J states after their argumentation dialogue.

The issue of the dispute will depend fully on the comparison between the strength of the considered arguments. In our example, that means that because the strength of W's argument ($Weight(w) = 4.2$) for going through the left road is stronger than the strength of J's argument ($Weight(j) = 4$) for going by the right road, J will concede. The social commitment proposed by W will be accepted and the one advocated by J rejected.

J: Ok, we will go through the left way.[15]

4.4 Argument integration

Here, we make the hypothesis that each agent fully integrates the other's point of view in his own cognitive coherence calculus. This means that the perceptions

[15] Concretely, this means that J's embedded request will be refused by W and W's offer finally accepted by J. All the opened games will thus be closed.

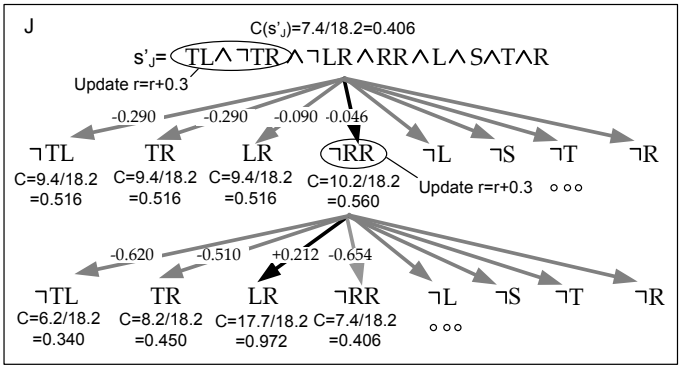

Fig. 5. J's reasoning from the state s'_J, resulting from the argumentation dialogue. Notice the attitude change.

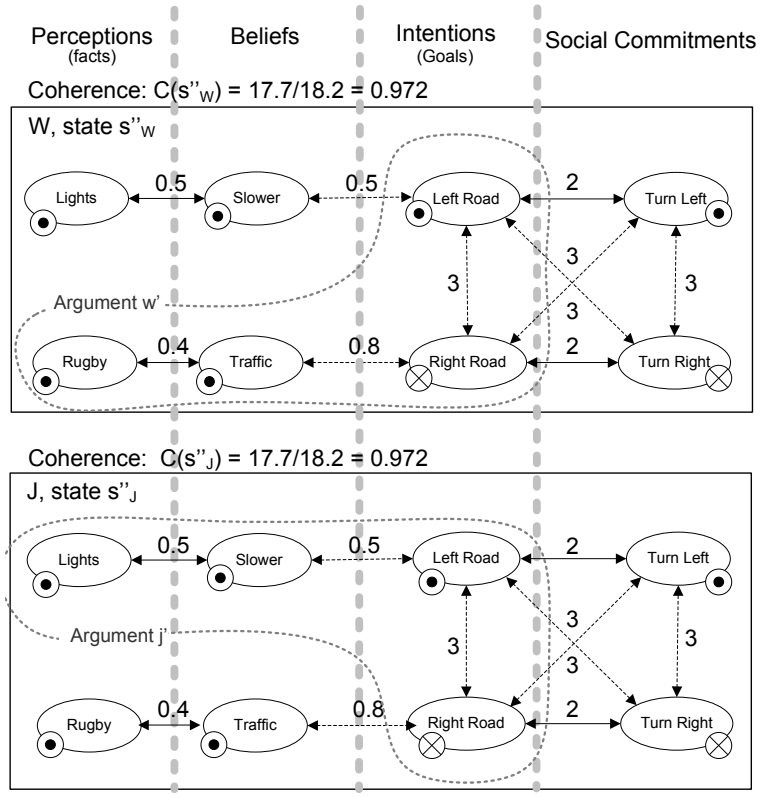

Fig. 6. Final states (after integration) for W and J.

and beliefs as well as goals and social commitments supporting the other's point of view are integrated in the cognitive model of the agent regardless to their strength. This corresponds to a fully cooperative and trustful cognitive behavior. Many other integration strategies are possible and will be discussed and compared as part of our future work.

Cooperation in cognitive coherence theory results from the fact that once an agent is aware (even partially) about the other's cognitive constraints, he will be able to take them into account in his own coherence seeking. This argument integration procedure is fully cooperative since the others' arguments will be fully taken into account in future reasoning. In the current model integration is done after the argument evaluation, thus being a post-evaluation memorization of arguments. Note that different choices may have been possible that will be inquired in future work.

In our example, argument evaluation and integration result in the cognitive models depicted by Figure 4. While W cannot improve his cognitive coherence anymore, Figure 5 shows J's reasoning which embed an attitude change. Figure 6 presents the final state of the agents which is an equilibrium (no element acceptance change can improve cognitive coherence). Notice that the agent coherence is not maximal (i.e. 1) because of the integration of J's argument which is against the chosen issue (and is valuable).

Finally, it is probable that W will turn left in order to fulfill the corresponding social commitment and advance the state of the environment...

5 Discussion

5.1 Comparison with Dung's approach to argumentation

If we represent our example of Figure 2 within the classical argumentation approach defined in [9], in which we call J's argument j and W's one w, we obtain the following argumentation framework: $< \{w, j\}, \{(w, j), (j, w)\} >$, composed of the two arguments and their attack relation. This particular argumentation framework has two *acceptable stable preferred extensions* (namely $\{w\}$ and $\{j\}$), which doesn't say much about persuasion. According to the semantics of acceptability in Dung's and subsequent approaches, a credulous agent accepts all acceptable extensions while a sceptical one only accepts the intersection of all acceptable extensions (which is void here). In other words, as noted in [3], Dung's approach to argumentation does not allow to fully treat persuasion.

In a multi-agent setting, preferences are needed in order to conclude (as shown by Amgoud and al. [1]). In our approach, preferences are implicit and follow from the coherence principle that coherence is preferred to incoherence. Since this is true both at the qualitative and quantitative levels, we don't need any extra treatment for taking preferences into accounts.

5.2 On bipolarity in the cognitive coherence approach

While Dung's framework only considers one type of interaction between arguments (i.e. attacks), it has been extended to take into account bipolarity, that

is the fact that supportive and negative arguments may be differentiated, which has been shown to be useful in a number of applications [2].

In our framework, the notion of argument can be refined to consider supportive argument as well as negative argument. Here, we provide the following definitions:

Definition 9. (Supportive Argument) *A **supportive argument** for an element acceptance (resp. rejection) is (1) an argument in the sense of definition 6 that is (2) optimally coherent with the acceptance (resp. rejection) of the conclusion.*

Definition 10. (Negative Argument) *A **negative argument** for an element acceptance (resp. rejection) is (1) an argument in the sense of definition 6 for which (2) there exist an assignation that would be more coherent than the current one in which the conclusion is rejected (resp. accepted).*

Using those definitions, it is possible to say that at the end of the dialogue (Figure 6):

- w' is a supportive argument for the acceptance of the intention to go by the left road (noted LR).
- while j' is a negative argument for the acceptance of LR.

Further relation(s) with previous work and other approaches to argumentation are let as future work.

5.3 Coverage of the presented approach

Our approach allows to cover a variety of argumentation dialogues. For example, argumentations that rely on elements types (cognitions types and their related resistance to change). For example, the following dialogue involves perception as an argument:

W: Google can answer a request in less than 2 seconds and gives you pertinent pages out of several millions ones.
J: No!
W: Yes.
J: How do you know?
W: I have seen it.

Also, while *social arguments* have not been considered in the literature yet, we think they are crucial in multi-agents settings. Here is an example, that can be captured by our approach, where J justify his decision using a social argument:

Q: Do you want to go to the cinema tonight?
J: No, I can't.
Q: Why?
J: I promise my boss to finish a paper tonight.

More generally, the treatment of the cognitive aspects of pragmatics model the persuasion process that allow to capture a variety of persuasive dialogues including those that does not involve argumentation. Here is an example of such dialogue:

Boss: You have to finish that paper tonight.
J: Yes.

In DIAGAL [18], an order given by an agent that has authority over an other one results in a social commitment being accepted by definition. However, J's behavior will still be guided by his coherence calculus and J will either enter an attitude change and accept the corresponding intention or cancel or violate this social commitment while coping the sanctions (which are taken into account in the agent reasoning through the resistance to change of the accepted commitment).

This shows how our approach integrates argumentation with other agent communication behavior through the modelling of the cognitive aspect of pragmatics that emphasizes the persuasive dimension of every communication. The limit case of argumentation dialogue being the one in which each argument consists of a single element, our approach can be seen as an attempt to unify argumentation-based frameworks with previous agent communication frameworks (specifically social commitment based communication) through some higher level concepts from cognitive sciences.

6 Conclusion

In this paper, we have highlighted the persuasive aspects inherent to every communication (thus including argumentation) by providing a model in which the cognitive response to persuasive message was modelled (by reifying the concept of attitude change when necessary). The strength of the proposed approach resides in the facts that: (1) all the steps of argumentation are computed using a single set of measures, i.e. the cognitive coherence metrics, (2) the approach is grounded in behavioral cognitive sciences rather than in dialectics and is part of a more general theory of mind, which covers many dimensions of the cognitive aspects of pragmatics and (3) our characterization is computational.

The presented framework has been developed in order to fill the need (that is not covered by previous approaches) of implementable argumentation based framework that are integrated to a more general agent architecture and communication framework. While promising, this alternative approach to argumentation requires more work. In particular, studying more precisely how this framework differs from and complements previous (dialectic based) proposals is in our future work list.

References

1. L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the 14th Conference on Uncertainty in Artifi-*

cial Intelligence (UAI 1998), pages 1–7, San Francisco CA, USA, 1998. Morgan Kaufmann Publishers.
2. L. Amgoud, C. Cayrol, and M.-C. Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In *10th International Workshop on Non-Monotonic Reasoning (NMR-2004)*, pages 1–9, 2004.
3. T.J.M Bench-Capon. Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.
4. J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. Van der Torre. The BOID architecture: Conflicts between beliefs, obligations, intention and desires. In *Proceedings of the Fifth International Conference on Autonomous Agent*, pages 9–16. ACM Press, 2001.
5. B. Chaib-draa, M. Bergeron, M.-A. Labrie, and P. Pasquier. Diagal: An agent communication language based on dialogue games and sustained by social commitments. *Journal of Autonomous agents and Multi-agents Systems (to appear)*, 2005.
6. ASPIC Consortium. Review on argumentation technology: State of the art, technical and user requirements. Prepared for the european commission, ASPIC(Argumentation Service Platform with Integrated Components), http://www.argumentation.org/, 2004.
7. ASPIC Consortium. Theoretical framework for argumentation. Prepared for the european commission, ASPIC(Argumentation Service Platform with Integrated Components), http://www.argumentation.org/, 2004.
8. F. Dehais and P. Pasquier. Approche Générique du Conflit. In D.L. Scapin and E. Vergisson, editors, *Ergonomie et Interaction Homme-Machine (ErgoIHM 2000)*, pages 56–63, France, 2000. ESTIA (École Supérieure des Technologies Industrielles Avancées).
9. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.
10. P. Erwin. *Attitudes and Persuasion*. Psychology Press, 2001.
11. L. Festinger. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957.
12. A. G. Greenwald. *Psychological Foundations of Attitude Change*, chapter Cognitive Learning, Cognitive Response to Persuasion and Attitude Change, pages 147–170. Academic Press, New York, 1968.
13. H. Jung and M. Tambe. Toward argumentation as distributed constraint satisfaction. In *Proceedings of the AAAI Fall Symposium on Negotiation Methods fos Autonomous Cooperative Systems*, 2001.
14. H. Jung, M. Tambe, and S. Kulkarni. Argumentation as distributed constraint satisfaction: Applications and results. In *Proceedings of the International Conference on Autonomous Agents (Agents'01)*, pages 324–331, Montreal, Canada, 2001. ACM Press.
15. B. Moulin, H. Irandoust, M. Blanger, and G. Desbordes. Explanation and argumentation capabilities: Towards the creation of more persuasive agents. *Artificial Intelligence Review*, 17(3):169–222, 2002.
16. P. Pasquier. *Aspects cognitifs des dialogues entre agents artificiels : l'approche par la cohérence cognitive*. PhD thesis, Laval University, Quebec, Canada, August 2005.
17. P. Pasquier, N. Andrillon, and B. Chaib-draa. An exploration in using cognitive coherence theory to automate BDI agents' communicational behavior. In F. Dignum, editor, *Advances in Agent Communication - International Workshop*

on Agent Communication Languages (ACL'03), volume 2922 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 37–58. Springer-Verlag, 2003.
18. P. Pasquier, M. Bergeron, and B. Chaib-draa. Diagal : a dialogue games based agent language. In *Proceedings of the XX European Conference of Artificial Intelligence (ECAI'04)*. Unpublished, 2004.
19. P. Pasquier, M. Bergeron, and B. Chaib-draa. DIAGAL: a Generic ACL for Open Systems. In *Proceedings of The Fifth International Workshop Engineering Societies in the Agents World (ESAW'04)*, volume 3451 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 139–152. Springer-Verlag, 2004.
20. P. Pasquier and B. Chaib-draa. The cognitive coherence approach for agent communication pragmatics. In *Proceedings of The Second International Joint Conference on Autonomous Agent and Multi-Agents Sytems (AAMAS'03)*, pages 544–552. ACM Press, 2003.
21. P. Pasquier and B. Chaib-draa. Agent communication pragmatics: The cognitive coherence approach. *Cognitive Systems*, 6(4):364–395, December 2005.
22. P. Pasquier, R. A. Flores, and B Chaib-draa. Modelling flexible social commitments and their enforcement. In *Proceedings of the Fifth International Workshop Engineering Societies in the Agents World (ESAW'04)*, volume 3451 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 153–165. Springer-Verlag, 2004.
23. I. Rahwan, S. Ramchurn, N. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation based negotiation. *Knowledge Engineering Review*, 18(4):343–375, 2003.
24. J-P. Sansonnet and E. Valencia. Dialogue between non-task oriented agents. In *Proceedings of the 4th Workshop on Agent Based Simulation (ABS'04)*, Montpelier, France, april 2003. http://www.limsi.fr/Individu/jps/research/buzz/buzz.htm.
25. M Sherif and C.I. Hovland. *Social Judgement*. Yale University Press, New Haven, USA, 1961.
26. R. Shultz and R. Lepper. *Cognitive Dissonance : progress in a pivotal theory in social psychology*, chapter Computer simulation of the cognitive dissonance reduction, pages 235–265. American Psychological Association, 1999.
27. M. P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7:97–113, 1999.
28. R. Sun. *Connectionist-Symbolic Integration*, chapter An introduction to hybrid connectionist-symbolic models. Lawrence Erlbaum Associates., 1997.
29. P. Thagard. *Coherence in Thought and Action*. The MIT Press: Cambridge, MA, USA, 2000.
30. P. Thagard. Probabilistic network and explanatory coherence. *Cognitive science Quaterly*, (1):91–114, 2000.
31. P. Thagard and K. Verbeurgt. Coherence as constraint satisfaction. *Cognitive Science*, 22:1–24, 1998.
32. F. H. van Eemeren and R. Grootendorst. *A Systematic Theory of Argumentation: the Pragma-Dialectical Approach*. Cambridge University Press, 2004.