Argumentation and Persuasion in the Cognitive Coherence Theory

Philippe Pasquier [a,1], Iyad Rahwan [b] Frank Dignum [c] Liz Sonenberg [a]
[a] *University of Melbourne, Australia*
[b] *British University of Dubai, UAE*
[c] *Utrecht University, The Netherlands*

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." [30, 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 [21,15,7,6]. 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 [18,19,16] 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

[1]Correspondence to: Philippe Pasquier. Tel.: +61 4 3341 2533; Fax: +61 3 9349 4596; E-mail: pasquier@unimelb.edu.au

cognitive response to persuasive arguments using a single set of principles (section 4). Finally, links with previous approaches are discussed (section **??**).

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 [27]. In that context, our main and original theoretical contribution has been to extend that model to communication (which has 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 have been made (logic-based [22], neural network or activation network based [24], probabilistic network [28], decision-theoretic, . . .), we present one that is constraint satisfaction based resulting in a simple symbolic-connexionist hybrid formalism (we refer the reader to [26] 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 came 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[1]. 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 permits the reification of 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:*

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

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 [29]. An agent can be partially defined as follows:

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

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

- *$\mathcal{P}$,$\mathcal{B}$,$\mathcal{I}$ 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 [20]). Social commitment has proven to be a powerful concept to capture the interdependencies between agents [25]. 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 thinks (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's 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 belong 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's preferences. In the cognitive coherence theory, according to the afore-mentioned coherence principle, coherence is preferred to incoherence which allows to define the following expected utility function[2].

Definition 6 (Expected Utility Function) *The expected utility for an agent to attempt to reach the state W' from the state W (which only differ by the acceptance state of a subset E of the agent's elements) is expressed as the difference between the incoherence before and after this change minus the cost of the dialogue moves (expressed in term of the resistance to change of the modified elements): $G(W') = C(W') - C(W) - \sum_{X \in E} Res(X)$.*

[2] Note that our expected utility function does not include any probabilities. This reflects the case of equiprobability in which the agent has no information about other's behavior. Notice that integrating algorithms to progressively learn such probabilities is an obvious perspective of the presented model.

At each step of his reasoning, an agent will search for a cognition acceptance state change which maximizes this expected utility. If this cognition 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.

A recursive version of the local search algorithm the agents use to maximize their cognitive coherence is presented in [16] 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. Update the resistance to change of the modified element to avoid looping.
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.

Since it 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 constraints that bind them[3]. We don't have a proof of correctness of this greedy algorithm in regards to the general coherence problem but, it behaved optimally 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.

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. This framework has been implemented and exemplified as presented and discussed in [17] and [19]. The presented practical framework relies on our dialogue games based agent communication language (DIAGAL) 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 researches 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 and roles, power relations, persuasive argumentation, material constraints, . . .) that an agent comes to ac-

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

cept 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.

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 mean for an end. the end being persuasion, that is attitude change. But at the same time, argumentation is a type of information disclosure and competitive (or malicious) agents can use this information to endorse non-cooperative behavior. In this paper, we won't address strategic issues related to argumentation.

In the cognitive coherence framework, argumentation will be used in a systematic way 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, i.e. element. Previous work has been made about argumentation as constraint propagation in the field of distributed constraint satisfaction [14].

Definition 7 ***(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 8 ***(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* $(H \cup \{h\}) \subset Arg_{\mathcal{I}\cup SC} \wedge (H \cup \{h\}) \cap SC \neq \varnothing$*;*

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) 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 indicate 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 [22] (inspired by cognitive coherence approach) has inquired this link further. For the sake of simplicity, in this paper, we will consider that the levels 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 [23], showed that each agent maintains 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 domains, 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 agents integrate the other's point of view in their own cognitive model 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 depends 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. We will also make the assumption that'all arguments are valid and meaningful.

There are two ways of dealing with evaluation and integration: (1) do a selective evaluation and then integration of arguments after some adjustment of strength due to the evaluation, (2) doing integration and seing evaluation as the side-effect consequence of the memorization of arguments by the agents. While in previous work [**?**] we have explored the first path, this paper exemplify the second one.

4.3. 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 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.

4.4. Argument evaluation

The main consequence of this integration procedure is that we don't need argument evaluation. Argument evaluation and eventual persuasion (attitude change) will be done by the cognitive coherence calculus as a result of the argument integration. According to our hypothesis, the behavior, decision or beliefs of the agent may be changed (or not) depending on the effect of the integration (i.e. memorization) of the argument encountered. This is quite intuitive according to our hypothesis of cooperation, sincerity and validity of the arguments.

5. Example

As an example, we consider the two agents W and J that are driving a car (shared resource) and have to decide which way to go next. Their initial states are represented by Figure 1,(a). Using the approach presented in this paper (on top of the one presented in previous work), they will generate the following dialogue (using DIAGAL dialogue games instead of natural language):

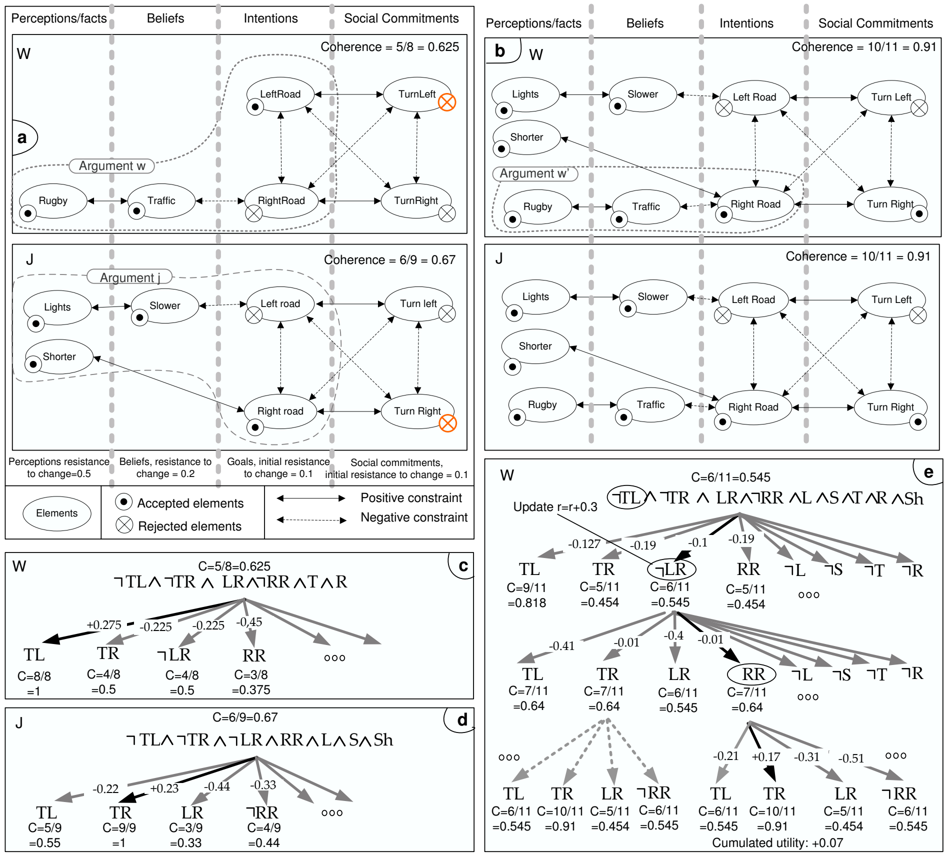

Figure 1. Parts (a) and (b) represents initial and final states for W and J, parts (c),(d) and (e) are traces of the local search algorithms, where arcs are labeled with the expected utility values. All the weights of the constraint are equal to 1.

J.1: I would turn right.
W.1: I would rather turn left. there is a rugby match and there gona be lots of traffic on the right road.
J.2: But, there 's a lot of lights on the left road, that will slow us down and the right road is shorter. Can't we turn right instead?
W.2: Ok, fine, lets turn right then.

Part d of Figure 1 indicates the coherence calculus that leads J to utter on offer (J.1) as an attempt to get the social commitment to turn right accepted. This explicit an external incoherence for W which cognitive coherence calculus leads to a different view, i.e. turning left (Part c of Figure 1). W than refuse J's offer, counter offer and disclose his arguments (W.1). That also makes explicit the external incoherence for J which counter argue (J.2). Part e of Figure 1, shows how W cognitive coherence calculus leads him to an attitude change and an acceptance of J's point of view after memorizing its argument (without forgetting his own). Finally, Part b of Figure 1 indicates the agents' states after that dialogue as well as the now shared and increased cognitive coherence.

6. Discussion

6.1. Comparison with Dung's approach to argumentation

If we represent our example of Figure 1 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 account.

6.2. 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 7 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 7 for which (2) there exist an assignation that would be more coherent than the current one in which the conclusion is rejected (resp. accepted).*

For example, in Figure 1): $argumentw$ is a supportive argument for the acceptance of the intention to go by the left road (noted LR), while w' is a negative argument for the acceptance of RR.

Because of the use of social commitments and the lack of links between social commitment and private cognition the integration part is usually not modeled in AI approach to argumentation issued from formal dialectics. Agents just do evaluation (of acceptable arguments), i.e. manage the public commitment store. We provide a more realistic model that includes links between social and private cognitions [**?**] and where the integration (i.e. memorization) of the others arguments is fully accounted. In that context evaluation, and more generally persuasion (i.e. eventual attitude change) is the result of the cognition process on the updated set of cognitive elements.

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

7. 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 frameworks that are integrated to a more general agent architecture and communication framework.

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 Artificial 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 non-monotonic 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. Bélanger, 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 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.

[19] P. Pasquier and B. Chaib-draa. Agent communication pragmatics: The cognitive coherence approach. *Cognitive Systems*, 6(4):364–395, December 2005.

[20] 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.

[21] I. Rahwan, S. Ramchurn, N. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation based negotiation. *Knowledge Engineering Review*, 18(4):343–375, 2003.

[22] 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.

[23] M Sherif and C.I. Hovland. *Social Judgement*. Yale University Press, New Haven, USA, 1961.

[24] 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.

[25] M. P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7:97–113, 1999.

[26] R. Sun. *Connectionist-Symbolic Integration*, chapter An introduction to hybrid connectionist-symbolic models. Lawrence Erlbaum Associates., 1997.

[27] P. Thagard. *Coherence in Thought and Action*. The MIT Press: Cambridge, MA, USA, 2000.

[28] P. Thagard. Probabilistic network and explanatory coherence. *Cognitive science Quaterly*, (1):91–114, 2000.

[29] P. Thagard and K. Verbeurgt. Coherence as constraint satisfaction. *Cognitive Science*, 22:1–24, 1998.

[30] F. H. van Eemeren and R. Grootendorst. *A Systematic Theory of Argumentation: the Pragma-Dialectical Approach*. Cambridge University Press, 2004.