Artificial Intelligence Research and Development
S. Sandri et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-061-2-224

Coherence as an Inclusive Notion of Rationality

Sindhu Joseph [1], Marco Schorlemmer and Carles Sierra
Artificial Intelligence Research Institute, IIIA
Spanish National Research Council, CSIC
Catalonia, Spain

Abstract. We propose a method based on coherence maximisation based on Thagard's theory of coherence to model rational behaviour. We show that the traditional analysis of behaviour based on utility functions can be emulated using our approach and prove that *the maximum element of a preference ordering is the same as that found by coherence maximisation over an automatically generated coherence graph.* In addition, it is easy and natural to model the dynamism and uncertainty of the beliefs of a rational agent with this approach. We illustrate the theorem by modelling the prisoner's dilemma.

Keywords. Utility, Coherence, Rationality

1. Introduction

In neo-classical economics, rationality is idealised to decisions that are optimal by maximising utility or profit, for realising goals of an adaptive system [5]. In this paper, we confine to this utility-maximising idealisation of rationality. A basic assumption while modelling a rational agent is the existence of an a priori ordering of preferences. Utility functions are practical representation mechanisms of preferences because it is possible to apply standard optimisation techniques given a utility function.

However, we encounter certain difficulties when using a utility function to model the behaviour of a rational agent. Firstly, an autonomous agent chooses to pursue an action by considering various factors. They may include the norms of a society it is part of, the reputation of the agent, the context of the action, altruism, etc. Those supporting a utility-based approach often claim that preferences are defined to include all these considerations. If this is the case, then such a preference is very hard to compute and we do not share the view that they are basic. On the contrary, preferences are the consequence of complex cognitive processes built upon basic cognitions.

The second difficulty is to determine the influence of a new information on the preference ordering. The preference ordering is either static or does not have a transparent computational mechanism to re-adjust to the new information. Both these difficulties arise mostly because there are no explicit methods to get to the preference ordering from

[1] Corresponding Author: IIIA-CSIC, Campus UAB, E-08193 Bellaterra (Barcelona), Catalonia, Spain; E-mail: joseph@iiia.csic.es.

the cognitions of an agent. In our view, preferences are indeed a consequence of the interaction of all the cognitions (goals, actions, plans, and beliefs), they are dynamic, possibly uncertain and imprecise. Therefore, a simplistic static linear ordering over possible actions, as a utility function establishes, falls short with respect to the requirement of autonomous agents.

In recent developments in autonomous agent theories, the theory of coherence offers a solution to the reasoning of individual and groups of agents [4,3]. The defining characteristic of such agents is that they take actions based on coherence maximisation. In this approach decision making is understood as a classification problem. That is, it attempts to put together those cognitions that strengthen each other while separating those that weaken or resist each other. In the context of coherence, preferences are interpreted as a consequence of this classification.

In this paper we show that, if preferences were basic, coherence function over a *utility coherence graph* could emulate the behaviour of a utility function, thus showing that the traditional interpretation of utility functions and of rationality can be preserved with our proposal. In particular, we prove that *the maximum element of a preference relation is exactly the same as that found by coherence maximisation over the corresponding utility coherence graph*. For this purpose we take the coherence framework introduced in Joseph et al. [3], which is based on Thagard's theory of coherence [6]. This framework provides a formal and a fully computational model of coherence. Further, this framework incorporates coherence maximisation into the cognitive representation of an agent. This makes it natural to accommodate new information. Moreover the framework defines a belief, desire, intention and norm coherence graphs separately [3] which makes it possible to embed a utility coherence graph (i.e. an intention coherence graph) in the global context of other cognitive coherence graphs. And a global coherence maximisation on such a graph allows an agent to consider other criteria such as norm abidance, values, or emotions.

The remainning paper is structured as follows. In Section 2, we briefly introduce the motivations behind Thagard's theory of coherence, and later introduce the basics of the coherence framework based on the work of Joseph et al. [3], which we use as the basis for our work. In Section 3 we prove that given a set of outcomes with preferences, there exists a coherence graph whose coherence maximising partition is the most preferred outcome, and we illustrate a coherence-driven agent taking rational (i.e., utility-maximising) decisions while only driven by coherence. The concluding remarks and a glimpse of the future work are in Section 4.

2. Coherence Theory and Framework

In this section, we introduce Thagard's theory of coherence, which is the major inspiration for this work. Then we briefly introduce a coherence-based framework [3], which is based on Thagard's theory. We make use of this framework to define our utility coherence graphs and simulate a rational agent.

2.1. Thagard's Coherence Theory

Thagard postulates that coherence theory is a cognitive theory with foundations in philosophy that approaches problems in terms of the satisfaction of multiple constraints within networks of highly interconnected elements [6,7]. At the interpretation level, Thagard's theory of coherence is the study of associations, that is, how a piece of information influences another and how best different pieces of information can be fitted together. In this regard, we can see each piece of information as imposing a constraint on another one, the constraint being positive or negative. A positive constraint strengthen pieces of information, thereby increasing coherence, while a negative constraint weaken them, thereby increasing incoherence. Hence, a coherence problem is to put together those pieces of information that have a positive constraint between them, while separating those having a negative constraint. If all pieces of information are classified in this manner, then all constraints are satisfied so that coherence is optimal. In general such an optimal partition does not exist, and hence we talk of maximising coherence.

As discussed in [6], this view of decision making is very different from those of classical decision making theories where the notion of *preference* is atomic and there is no conceptual understanding of how preferences can be formed. In contrast, coherence based decision making tries to understand and evaluate these preferences from the available complex network of constraints. The assumption here is more basic because the only knowledge available to us are the various interacting constraints between pieces of information. That is, with coherence maximisation, a highly desired state of the world (preferred in a classical sense) may get discarded in front of a less desired state of the world because it is incoherent with the rest of the beliefs, desires or intentions.

2.2. Coherence Framework

Since our aim is to model preferences and simulate rational decision making using coherence maximisation, in this section we summarise one of the computational frameworks for coherence introduced in Joseph et al. [3]. It differs from other coherence-driven approaches in extending agent theories [4] as it modifies the way an agent framework is perceived by making the associations in the cognitions explicit in representation and analysis. That is, in this framework coherence is treated as a fundamental property of the mind of an agent. Further, it is generic and fully computational. Hence, it suits well for our purposes, as we would like to have a framework that has a uniform representation and is computational. Further, we need to have a generic framework, as it would allow us to have no assumptions on the type of agents. In the following, we briefly introduce the necessary definitions of this framework which forms the base for the remainning sections. The intuition behind these definitions and a few examples are given in [3]. The core notion is that of a *coherence graph* whose nodes represent pieces of information and whose weighted edges represent the degree of coherence or incoherence between nodes.

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

1. V is a finite set of nodes representing pieces of information.
2. $E \subseteq \{\{v, w\} | v, w \in V\}$ is a finite set of edges representing the coherence or incoherence between pieces of information, and which we shall call *constraints*.
3. $\zeta : E \to \mathbb{R} \setminus 0$ is an edge-weighted function that assigns a negative or positive value to the coherence between pieces of information, and which we shall call *coherence function*.

Every coherence graph is associated with a number called the *coherence of the graph*. Based on Thagard's formalism, this can be calculated by partitionning the set of nodes V of the graph in two sets, $\mathcal{A}$ and $V \setminus \mathcal{A}$, where $\mathcal{A}$ contains the accepted elements of V, and $V \setminus \mathcal{A}$ contains the rejected ones. The aim is to partition V in such a way that the maximum number of constraints are satisfied, taking their value into account. A constraint is satisfied only if it is positive and both the end nodes are in the same set, or negative and the end nodes are in complementary sets. The following definitions help clarify this idea.

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

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

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

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

$$\sigma(g, \mathcal{A}) = \frac{\sum\limits_{\{v,w\} \in C_{\mathcal{A}}} | \zeta(\{v, w\}) |}{| E |} \ .$$

Notice that, by Definitions 2 and 3,

$$\sigma(g, \mathcal{A}) = \sigma(g, V \setminus \mathcal{A}) \ . \tag{1}$$

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

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

If for some partition $(\mathcal{A}, V \setminus \mathcal{A})$ of V, the coherence is maximal (i.e., $\kappa(g) = \sigma(g, \mathcal{A})$ and thus $\mathcal{A} = \arg\max_{\mathcal{A} \subseteq V} \sigma(g, \mathcal{A})$), then the set $\mathcal{A}$ is called the *accepted* set and $V \setminus \mathcal{A}$ the *rejected* set of the partition.

A *coherence-driven agent* is an agent that always chooses an action based on the maximisation of coherence of its cognitions, norms, and other social commitments. In a coherence-driven agent, the agent theories are expressed as coherence graphs, so that the cognitions are nodes of the graphs. In Joseph et al [3], the authors define different coherence graphs for beliefs, desires, intentions, and norms which they call belief, desire, intention and norm coherence graphs respectively. Hence for example, a belief cognition is a node in a belief coherence graph. Both these specific cognitive coherence graphs and mechanisms to combine them are defined in [3]. For the purpose of the paper, it is sufficient to understand that each agent finally generates a coherence graph which it partitions to maximise coherence. Based on the coherence maximisation it then accepts certain cognitions while rejecting certain others. Any update of the agent theory triggers the whole process of first updating the specific coherence graph affected by the update and then updating the combined graph all over again.

3. Utility Coherence Graphs

As discussed in the introduction, a rational agent is an agent that chooses an outcome that maximises the utility among a finite set of outcomes based on a preference relation on them. Here, as in utility maximisation, we assume that there is an a priori preference established by the agent. Based on this assumption, we prove that there exists a coherence graph that produces a utility maximising outcome by way of maximising coherence. We call such a graph a *utility coherence graph*. Before we discuss the theorem and its proof, we define certain preliminaries.

3.1. Preference Relation

Preferences are relevant when it is necessary to examine the behaviour of an individual who must choose from a set of outcomes O. A preference relation basically describes an order of preference of an agent among a set of alternative outcomes.

Definition 5 A *preference relation* $\succcurlyeq$ on a finite set of outcomes O is a total pre-order on O, i.e., for all $o, p, q \in O$,

- $o \succcurlyeq o$ ($\succcurlyeq$ is reflexive)
- if $o \succcurlyeq p$ and $p \succcurlyeq q$ then $o \succcurlyeq q$ ($\succcurlyeq$ is transitive)
- $o \succcurlyeq p$ or $p \succcurlyeq o$ ($\succcurlyeq$ is complete)

When $o \succcurlyeq p$ we say that *o is at least as preferable as p*. We write $o \sim p$ when $o \succcurlyeq p$ and $p \succcurlyeq o$, and $o \succ p$ when $o \succcurlyeq p$ but $o \not\preccurlyeq p$. The down-set of $o \in O$ is $\downarrow o = \{p \in O | o \succcurlyeq p\}$. Further, we shall require the set O to have at least three outcomes. With two or less outcomes, accepting the most preferable one is trivial and not suitable to be determined via coherence due to the property of σ given in Equation (1) (see Section 2.2). In this paper, we consider only those preference relations on O, that have a maximum.

3.2. A Utility Coherence Graph

As mentioned in the introduction, a utility function is used to assign a numerical value to the preference ordering, so that we can employ numerical techniques to select an outcome that is most preferred. The aim of this section is to prove that, given a preference ordering, there exists a coherence maximising function over a utility coherence graph which reproduces the behaviour of a utility maximising function. Further, we also show how such a graph would look like.

We now state a lemma that defines the necessary conditions under which a graph g will be a utility coherence graph. The first condition states that $o \succ p$ with respect to a preference relation, if and only if, their individual coherence strengths preserve the ordering. In other words, it is more coherent to accept a more preferred outcome to a less preferred one. The second condition states that, for the maximum in O, it is more coherent to accept the maximum alone than accepting it together with any other less preferred outcomes.

To prove the lemma, we define a coherence graph as follows. We take the set of outcomes as nodes of the graph. Since the outcomes are mutually exclusive, it is natural to assign negative coherences between them. The degrees of incoherence depend on the number of pairs of equally preferred outcomes, and to satisfy the second condition of the lemma below, the degrees should decrease exponentially as less preferred nodes are linked in the graph. Hence, we use the cardinalities of down-sets to define the degree of coherence between two nodes a and b as $-|O|^{|\downarrow a|+|\downarrow b|}$. However, there may exist other ways to determine the degrees of incoherence that satisfy the conditions of the lemma.

Lemma 1 *Let O be a finite set of outcomes such that $|O| \geq 3$, and let $\succcurlyeq$ be a preference relation on O. Then there exists a coherence graph g such that,*

a) *for all $o, p \in O$, $o \succ p$ if and only if $\sigma(g, \{o\}) > \sigma(g, \{p\})$;*
b) *if o is the maximum of O by $\succcurlyeq$, for all $P \subseteq O$ such that $o \in P$ and $P \neq \{o\}$; then $\sigma(g, \{o\}) > \sigma(g, P)$.*

PROOF: Let $g = \langle V, E, \zeta \rangle$ be a coherence graph such that

- $V = O$
- $E = \{\{o, p\} \mid o, p \in O\}$
- for all $o, p \in O \quad \zeta(\{o, p\}) = -|O|^{|\downarrow o|+|\downarrow p|}$ (Note that, for the purpose of not cluttering the equations in the proof, we henceforth denote $|\downarrow a|$ in the exponent simply as a).

That is, in g all nodes are incoherent between each other. In such a graph, given a partition $(\mathcal{A}, V \setminus \mathcal{A})$ of V, the set of satisfied constraints is, by Definition 2, $C_{\mathcal{A}} = \bigcup_{v \in \mathcal{A}} \bigcup_{w \in V \setminus \mathcal{A}} \{\{v, w\}\}$. Consequently, by Definition 3,

$$\sigma(g, \mathcal{A}) = \frac{\sum_{v \in \mathcal{A}} \sum_{w \in V \setminus \mathcal{A}} |\zeta(\{v, w\})|}{|E|} \; .$$

Since in g, $V = O$ and, for all $o, p \in O$, $|\zeta(\{o, p\})| = |O|^{o+p} = |O|^o \cdot |O|^p$, we have that

$$\sigma(g, \mathcal{A}) = \frac{\sum_{o \in \mathcal{A}} \sum_{p \in O \setminus \mathcal{A}} |O|^o \cdot |O|^p}{|E|} \; .$$

a) For all $o, p \in O$, $\sigma(g, \{o\}) > \sigma(g, \{p\})$ if and only if

$$\sum_{q \in O \setminus \{o\}} |O|^o \cdot |O|^q > \sum_{q \in O \setminus \{p\}} |O|^p \cdot |O|^q \; .$$

Splitting the summations to extract the common term $|O|^o \cdot |O|^p$ and cancelling it out, the inequality is equivalent to

$$|O|^{o+p} + |O|^{o} \cdot \sum_{q \in O \setminus \{o,p\}} |O|^{q} \quad > \quad |O|^{p+o} + |O|^{p} \cdot \sum_{q \in O \setminus \{o,p\}} |O|^{q} \ ,$$

which is equivalent to

$$|O|^{o} \cdot \sum_{q \in O \setminus \{o,p\}} |O|^{q} \quad > \quad |O|^{p} \cdot \sum_{q \in O \setminus \{o,p\}} |O|^{q} \ .$$

Since $|O| \geq 3$, we have that $\sum_{q \in O \setminus \{o,p\}} |O|^{q} > 0$, and consequently, the above inequality is equivalent to $|O|^{o} > |O|^{p}$, which holds if and only if $|\downarrow o| > |\downarrow p|$, and if and only if $o \succ p$.

b) Let o be the maximum in O, and let $P \subseteq O$ such that $o \in P$ and $P \neq \{o\}$. By Definition 3,

$$\sigma(g, \{o\}) > \sigma(g, P) \text{ if and only if } \sum_{q \in O \setminus \{o\}} |O|^{o} \cdot |O|^{q} \quad > \quad \sum_{p \in P} |O|^{p} \cdot \sum_{q \in O \setminus P} |O|^{q} \ .$$

Taking the common terms from both the sides, we get,

$$|O|^{o} \cdot \Big(\sum_{p \in P \setminus \{o\}} |O|^{p} + \sum_{q \in O \setminus P} |O|^{q} \Big) \quad > \quad |O|^{o} \cdot \sum_{q \in O \setminus P} |O|^{q} + \sum_{p \in P \setminus \{o\}} |O|^{p} \cdot \sum_{q \in O \setminus P} |O|^{q} \ .$$

Cancelling the common term $|O|^{o} \cdot \sum_{q \in O \setminus P} |O|^{q}$ on both the sides, we have

$$|O|^{o} \cdot \sum_{p \in P \setminus \{o\}} |O|^{p} \quad > \quad \sum_{p \in P \setminus \{o\}} |O|^{p} \cdot \sum_{q \in O \setminus P} |O|^{q} \ .$$

Since $\sum_{p \in P \setminus \{o\}} |O|^{p} > 0$, the above inequality is equivalent to

$$|O|^{o} \quad > \quad \sum_{q \in O \setminus P} |O|^{q} \ .$$

To prove this inequality, let $p \in P$ and $p \neq o$. Since $O \setminus P \subseteq O \setminus \{o, p\}$, we have

$$\sum_{q \in O \setminus P} |O|^{q} \quad \leq \quad \sum_{q \in O \setminus \{o,p\}} |O|^{q} \ .$$

Since o is the maximum, for all $q \in O$ and $q \neq o$, $|O|^{q} \leq |O|^{o-1}$. Consequently,

$$\sum_{q \in O \setminus \{o,p\}} |O|^{q} \leq (|O| - 2) \cdot |O|^{o-1} < |O| \cdot |O|^{o-1} = |O|^{o} \ .$$

□

Definition 6 A coherence graph g is a utility coherence graph if and only if it satisfies the conditions of Lemma 1.

	C_x	D_x
C_y	3,3	0,5
D_y	5,0	1,1

Table 1. Outcomes in the Prisoner's Dilemma

We now define the theorem that may be summarised by saying that, there exists a coherence graph with certain properties that respects an a priori preference ordering. The theorem makes it possible to reformulate preference orderings as utility coherence graphs and to compute the maximally preferred outcome via a maximisation process over the graphs.

Theorem 1 *Given a finite set of outcomes O such that $|O| \geq 3$, and a preference relation $\succeq$ on O, there exists a coherence graph g such that, o is the maximum in O with respect to $\succeq$ if and only if $\{o\}$ is the accepted set of a coherence maximising partition of g.*

PROOF: We proceed by contradiction. First let us assume that o is not the maximum element of O and that $\{o\} = \arg\max_{Q \subseteq O} \sigma(g, Q)$. However, if o is not the maximum in O, then there exists a $p \in O$ such that $p \succ o$. Then, by Lemma 1 a), we have $\sigma(g, \{p\}) > \sigma(g, \{o\})$, which is a contradiction.

Now let us assume that o is the maximum in O and $\{o\} \neq \arg\max_{Q \subseteq O} \sigma(g, Q)$. Then, there exists a $P \subseteq O$ such that $P = \arg\max_{Q \subseteq O} \sigma(g, Q)$ and both $P \neq \{o\}$ and $O \setminus P \neq \{o\}$, because of the property of σ given in Equation (1) stated in Section 2.2.

On the one hand, if $o \in P$, then, by Lemma 1 b), we have $\sigma(g, \{o\}) > \sigma(g, P)$, which is a contradiction. On the other hand, if $o \notin P$, then $o \in O \setminus P$. By Lemma 1 b), $\sigma(g, \{o\}) > \sigma(g, O \setminus P)$. Then, b Equation (1), $\sigma(g, \{o\}) > \sigma(g, P)$, which is a contradiction. □

In a prisoner's dilemma game between two players X and Y, player X has the preference ordering of the outcomes as $(D_X, C_Y) \succ (C_X, C_Y) \succ (D_X, D_Y) \succ (C_X, D_Y)$, where (D_X, C_Y) stands for *defect move by player X and co-operate move by player Y*. We here do not represent the outcomes as such, but the beliefs of player X on the possible outcomes.[2] So our nodes are of the form $B_X(D_X, C_Y)$ rather than (D_X, C_Y). Hence the graph we consider is a belief coherence graph. The scores of the outcomes are as shown in Figure 1. Note that we here only model one of the prisoners by giving his/her preference ordering.

We model the preferences of player X using the utility coherence graph depicted in Figure 1. The nodes in the graph are beliefs about the outcomes. Since there are four outcomes in this example, we take 4 as the base value for the coherence degrees. Then, for instance, the coherence value between $B_x(D_X, C_Y)$ and $B_x(C_X, C_Y)$ is $-4^{|\downarrow B_x(D_x, C_y)| + |\downarrow B_x(C_x, C_y)|}$ and is equal to $4^{(4+3)}$. We can see that the accepted set of the coherence-maximising partition (Figure 1) is the singleton set $\{B_X(D_X, C_Y)\}$,

[2] The outcome is the result of the decision of both agents, and therefore which one will end up being selected is a matter of belief for an agent as it is not in control of the actions of the other agent.

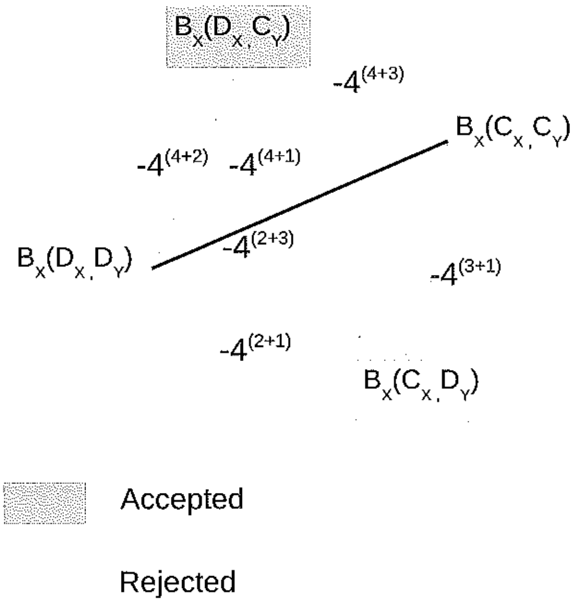

Figure 1. Preference ordering expressed as a utility coherence graph

which is the same as the most preferred outcome. Thus, we show that a coherence maximising function emulates the behaviour of a utility maximising function.

4. Discussion and Future Work

We have modelled preferences using the notion of utility coherence graphs. By proving the corresponding theorem, we say the case is indeed possible when preferences have a maximum. Utility coherence graphs, unlike utility functions, have the advantage of being integrated into the representation of a rational agent, allow for dynamic handling of new information affecting the preferences, help maintain the autonomy of the agent, and permit the representation of uncertainty about cognitions.

Though there are no closely following related work, we took inspiration from the line of thought expressed by Herbert Simon [5]. He distinguishes operational rationality from normative rationality due to limits on human intelligence. We identify with his view that satisficing has to be replaced by optimality, which is more a theoretical possibility than a realistic alternative. Here we propose a computationally feasible model, which subsumes the utility-maximising interpretation of rationality and yet has possibilities to broaden its limits.

Since this is the first step in the direction of attempting to contribute to alternative models for rationality that are more realistic, there are a couple of future work which are immediate. In most of game theory, expected utility rather than utility is used to model rationality in uncertain circumstances. In the future work, we plan to explicitly deal with expected utilities, however, we expect it to be already taken care of in our coherence based model as is evident from the example. That is, we are guessing that probability of an outcome can be represented by a set of beliefs and their interactions, which is taken care of in our coherence-driven agents. Another important advancement of this line of thought would be to define analogous notions to Nash Equilibrium and Pareto Optimality.

References

[1] Jan Broersen and Mehdi Dastani and Joris Hulstijn and Zisheng Huang and Leendert van der Torre, *The BOID architecture: conflicts between beliefs, obligations, intentions and desires*, AGENTS '01: Proceedings of the fifth international conference on Autonomous agents, 9–16, ACM, 2001,

[2] Sindhu Joseph and Carles Sierra and Marco Schorlemmer, *A coherence based framework for institutional agents*, Lecture Notes in Computer Science, volume 4870, year 2007,

[3] Sindhu Joseph and Carles Sierra and Marco Schorlemmer and Pilar Dellunde, *Formalising deductive coherence: An application to norm evaluation*, Normas'08(Extended version), Technical Report(RR-IIIA-2008-02), note http://www.iiia.csic.es/ sierra/papers/2009/Coherence.pdf, year 2009,

[4] Philippe Pasquier and Brahim Chaib-draa, *The cognitive coherence approach for agent communication pragmatics*, Second International Joint Conference on Autonomous Agents and Multiagent Systems, 2003,

[5] Herbert A. Simon, 1969, *The Sciences of the Artificial*, MIT Press,

[6] Paul Thagard, *Coherence in Thought and Action*, MIT Press, 2002,

[7] Paul Thagard, *Hot Thought*, MIT Press, 2006,