# Changing One's Mind: Erase or Rewind? Possibilistic Belief Revision with Fuzzy Argumentation based on Trust

Célia da Costa Pereira, Andrea G.B. Tettamanzi, Serena Villata

# Changing One's Mind: Erase or Rewind? Possibilistic Belief Revision with Fuzzy Argumentation based on Trust

**Célia da Costa Pereira**
Université de Nice Sophia Antipolis, Lab. I3S, France
celia.pereira@unice.fr

**Andrea G. B. Tettamanzi**
Università degli Studi di Milano DTI, Italy
andrea.tettamanzi@unimi.it

**Serena Villata** *
INRIA Sophia Antipolis, France
serena.villata@inria.fr

## Abstract

We address the issue, in cognitive agents, of possible loss of previous information, which later might turn out to be correct when new information becomes available. To this aim, we propose a framework for changing the agent's mind without erasing forever previous information, thus allowing its recovery in case the change turns out to be wrong. In this new framework, a piece of information is represented as an argument which can be more or less accepted depending on the trustworthiness of the agent who proposes it. We adopt possibility theory to represent uncertainty about the information, and to model the fact that information sources can be only partially trusted. The originality of the proposed framework lies in the following two points: (i) argument reinstatement is mirrored in belief reinstatement in order to avoid the loss of previous information; (ii) new incoming information is represented under the form of arguments and it is associated with a plausibility degree depending on the trustworthiness of the information source.

## 1 Introduction and related work

In a multiagent environment, belief revision aims at describing the changes in the agents mind in response to new information. On the other hand, one of the important concerns in argumentation is the strategies employed by an agent in order to succeed in changing the mind of another agent. To this aim, the agent must provide *good enough* reasons to (justify and then) succeed in such request of change. We can then view argumentation as an "incitement" to make an agent change its mind.

Despite the existence of such a clear complementarity between these two fields of Artificial Intelligence, there are few works integrating them in a unitary multiagent framework. However, a consensus exists on the opportunity of integrating belief revision and argumentation. [Cayrol *et al.*, 2010] do not integrate belief revision and argumentation, but propose a work on "revision in argumentation frameworks" in which they transpose the basic issue of revision into argumentation theory. They study the impact of the arrival of a new argument on the set of extensions of an abstract argumentation framework. [Quignard and Baker, 1997] describe a model for argumentation in agent interaction that shows how belief conflicts may be resolved by considering the relations between the agents' cognitive states and their choice of relevant argumentation moves. [Paglieri and Castelfranchi, 2006] claim that belief revision and argumentation can be seen as, respectively, the cognitive and social sides of the same epistemic coin and propose a preliminary framework which follows Toulmin's layout of argumentation. [Falappa *et al.*, 2009] survey relevant work combining belief revision and argumentation. Besides, they develop a conceptual view on argumentation and belief revision as complementary disciplines used to explain reasoning. They propose four basic steps of reasoning in multiagent systems.

- *Receiving new information*: new information can be represented as a propositional fact provided with a degree of plausibility;
- *evaluating new information*: the origin of new information decisively influences the agent's willingness to adopt it;
- *changing beliefs*: the agent uses belief revision techniques to change its epistemic state according to the new adopted information;
- *inference:* the agent's behavior is influenced by the most plausible beliefs resulting from its new epistemic state.

This paper is not "just" about integrating belief revision and argumentation in a single framework. It aims at using the strength resulting from such a combination to solve the problem of loss of information in the case of reinstatement of previous information in multiagent systems. More precisely, we answer the question "in case of such a reinstatement, *how* to recover from the loss of previous information which should become acceptable with new information, and to which extent old information should be recovered?"

The proposed framework integrates the first three basic steps considered by Falappa and colleagues. Indeed, in order to represent real situations more faithfully, we consider that new information is associated with a degree of plausibility which represents the trustworthiness, for the agent, of the source of information. This is in line with some work in

*The third author acknowledges support of the DataLift Project ANR-10-CORD-09 founded by the French National Research Agency.

the literature, like, for example, [da Costa Pereira and Tettamanzi, 2010], but the originality, which is the main difference with the previously cited authors, lies in the fact that a piece of information is represented as an argument which can be more or less acceptable. Therefore, such a degree directly influences the evaluation, performed by an agent, of new information and, as a consequence, it also influences the extent to which an agent changes its mind. Based on these considerations, we propose a fuzzy reinstatement algorithm which provides a satisfactory answer to our research question, which may be broken down into the following subquestions:

- How to represent arguments and beliefs in this setting?
- How to define a fuzzy evaluation of the arguments?
- How to address the change in the agent's cognitive state?

The first step is about determining the most suitable representation of partially trusted arguments and beliefs. Arguments, of the form $\langle \Phi, \phi \rangle$, support the agents' beliefs, which can be represented as the conclusions of structured arguments. The trustworthiness of a source can be measured by using probabilities only in the case in which data are available based on past experiences, for example. In many realistic cases, such data is not available. It is well known that possibilistic logic is well suited to deal with incomplete information. For example, [Amgoud and Prade, 2004] introduce a unified negotiation framework based on possibilistic logic to represent the agent's beliefs, preferences, decision, and revision under an argumentation point of view. A fuzzy labeling will then determine the fuzzy set of the agent's beliefs. [da Costa Pereira and Tettamanzi, 2010] adopt the representation of uncertain beliefs proposed in [Dubois and Prade, 1997]. The main point of their proposal may be described as belonging to the fourth among the basic steps proposed by [Falappa *et al.*, 2009], in the sense that they derive the most useful goals from the most plausible beliefs and desires. However, their approach for representing the changes in the agent's beliefs is not argumentation-based and cannot treat reinstatement in a satisfactory way.

The second step is about defining an algorithm which allows a fuzzy evaluation of the arguments. In crisp argumentation, arguments are evaluated, following a specific semantics, as acceptable or not acceptable, as shown by [Dung, 1995]. Intuitively, accepted arguments are those arguments which are not attacked by other accepted arguments and unaccepted arguments are those attacked by accepted arguments. Given an accepted argument, its conclusion can be adopted as belief in the agent's belief base. To represent the degrees of trust, we rethink the usual crisp argument evaluation [Dung, 1995; Caminada, 2006] by evaluating arguments in terms of fuzzy degrees of acceptability.

The third step is the choice about how to address the change in the cognitive state of the agent. As observed by [Dragoni and Giorgini, 1996] and [Delgrande *et al.*, 2006], for example, the main approaches to belief revision adopt the principle of the "priority to incoming information" but, in the context of multiagent systems, this principle presents some drawbacks. In particular, in a static situation, the chronological sequence of arrival of distinct pieces of information has nothing to do with their trustability or importance. This is supported also by [Gabbay *et al.*, 2003], where revision algorithms are presented in order to take into account the history of previous revisions as well as possible revision options which were first discarded but may now be pursued. The assumption we put forward in this paper is that, even if the agent accepts the incoming information throwing away part of the previously adopted belief base, this change must not be irrevocable. This means that, in the future, new information may turn the tide in such a way to have the past incoming information excluded from the belief-base and the original belief *somehow* reintegrated. This is exactly what happens in argumentation under the name of *reinstatement* principle. The difference, which is also one of the original contributions of this paper, is that the extent of the integration depends on the agent's trust in the source. Indeed, we evaluate arguments in a gradual way depending on such a degree of trust.

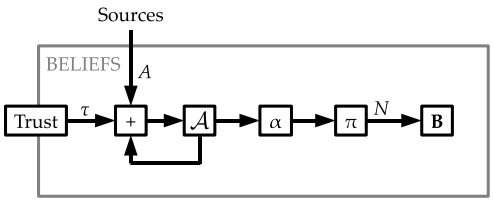


Figure 1: A schematic illustration of the proposed framework.

A schematic illustration of the proposed framework is visualized in Figure 1. The framework may be regarded as a belief revision model, based on argumentation. An agent interacts with the world by receiving arguments $A$ from one or more *sources*. The agent's internal mental state is completely described by a fuzzy set of trustful arguments $\mathcal{A}$, from which the beliefs of the agent may be derived. A *trust* module, whose details are not covered in this paper, assigns a trust degree $\tau$ to each source. As new arguments $A$ are received, they are added to $\mathcal{A}$ with the same membership degree as the degree $\tau$ to which their source is trusted. Fuzzy labeling of $\mathcal{A}$ yields a fuzzy reinstatement labeling $\alpha$, which may be regarded as a fuzzy set of acceptable arguments, whose consequences induce a possibility distribution $\pi$, from which an explicit representation $\mathbf{B}$ of the agent's beliefs is constructed as the necessity measure $N$ of possibility distribution $\pi$. Notice that we do not make any further assumptions on the trust model. This is out of the scope of this paper.

The paper is organized as follows: Section 2 provides the basic concepts of argumentation theory, Section 3 introduces the fuzzy evaluation of arguments, Section 4 discusses how beliefs change with respect to the acceptability degree of the arguments. Finally, some conclusions are drawn.

## 2 Argumentation theory and labeling

We provide the basic concepts and insights of Dung's abstract argumentation [Dung, 1995].

**Definition 1** *(Abstract argumentation framework) An abstract argumentation framework is a pair $\langle \mathcal{A}, \rightarrow \rangle$ where $\mathcal{A}$ is a set of elements called arguments and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation called attack. We say that an argument $A_i$*

*attacks an argument $A_j$ if and only if $(A_i, A_j) \in \rightarrow$.*

[Dung, 1995] presents several acceptability semantics which produce zero, one, or several sets of accepted arguments. These semantics are grounded on two main concepts, called conflict-freeness and defence.

**Definition 2** *(Conflict-free, Defence) Let $C \subseteq \mathcal{A}$. A set $C$ is conflict-free if and only if there exist no $A_i, A_j \in C$ such that $A_i \rightarrow A_j$. A set $C$ defends an argument $A_i$ if and only if for each argument $A_j \in A$ if $A_j$ attacks $A_i$ then there exists $A_k \in C$ such that $A_k$ attacks $A_j$.*

**Definition 3** *(Acceptability semantics) Let $C$ be a conflict-free set of arguments, and let $\mathcal{D} : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ be a function such that $\mathcal{D}(C) = \{A | C\ defends\ A\}$.*

- *$C$ is admissible if and only if $C \subseteq \mathcal{D}(C)$.*
- *$C$ is a complete extension if and only if $C = \mathcal{D}(C)$.*
- *$C$ is a grounded extension if and only if it is the smallest (w.r.t. set inclusion) complete extension.*
- *$C$ is a preferred extension if and only if it is a maximal (w.r.t. set inclusion) complete extension.*
- *$C$ is a stable extension if and only if it is a preferred extension that attacks all arguments in $\mathcal{A} \setminus C$.*

The concepts of Dung's semantics are originally stated in terms of sets of arguments. It is equal to express these concepts using argument *labeling*. This approach has been proposed firstly by [Jakobovits and Vermeir, 1999] and [Verheij, 2003], and then further developed by [Caminada, 2006] with the aim of providing quality postulates for dealing with the reinstatement of arguments. Given that $A_1 \rightarrow A_2$ and $A_2 \rightarrow A_3$, we have that argument $A_1$ reinstates argument $A_3$, i.e., it makes argument $A_3$ accepted by attacking the attacker of $A_3$. In a reinstatement labeling [Caminada, 2006], an argument is labeled "in" if all its attackers are labeled "out" and it is labeled "out" if it has at least an attacker which is labeled "in".

**Definition 4** *(AF-labeling [Caminada, 2006]) Let $\langle \mathcal{A}, \rightarrow \rangle$ be an abstract argumentation framework. An AF-labeling is a total function $lab : \mathcal{A} \rightarrow \{in,\ out,\ undec\}$. We define $in(lab) = \{A_i \in \mathcal{A} | lab(A_i) = in\}$, $out(lab) = \{A_i \in \mathcal{A} | lab(A_i) = out\}$, $undec(lab) = \{A_i \in \mathcal{A} | lab(A_i) = undec\}$.*

**Definition 5** *(Reinstatement labeling [Caminada, 2006]) Let $lab$ be an AF-labeling. We say that $lab$ is a reinstatement labeling iff it satisfies the following:*

- *$\forall A_i \in \mathcal{A} : (lab(A_i) = out \equiv \exists A_j \in \mathcal{A} : (A_j \rightarrow A_i \wedge lab(A_j) = in))$ and*
- *$\forall A_i \in \mathcal{A} : (lab(A_i) = in \equiv \forall A_j \in \mathcal{A} : (A_j \rightarrow A_i \supset lab(A_j) = out))$ and*
- *$\forall A_i \in \mathcal{A} : (lab(A_i) = undec \equiv \exists A_j \in \mathcal{A} : (A_j \rightarrow A_i \wedge \neg(lab(A_j) = out)) \wedge \nexists A_k \in \mathcal{A} : (A_k \rightarrow A_i \wedge lab(A_k) = in)$.*

*A reinstatement labeling is called illegal if the above conditions do not hold.*

A classical propositional language may be used to represent information for manipulation by a cognitive agent.

**Definition 6** *(Language) Let $\mathrm{Prop}$ be a* finite [1] *set of atomic propositions and let $\mathcal{L}$ be the propositional language such that $\mathrm{Prop} \cup \{\top, \bot\} \subseteq \mathcal{L}$, and, $\forall \phi, \psi \in \mathcal{L}$, $\neg\phi \in \mathcal{L}$, $\phi \wedge \psi \in \mathcal{L}$, $\phi \vee \psi \in \mathcal{L}$.*

As usual, one may define additional logical connectives and consider them as useful shorthands for combinations of connectives of $\mathcal{L}$, e.g., $\phi \supset \psi \equiv \neg\phi \vee \psi$.

We will denote by $\Omega = \{0,1\}^{\mathrm{Prop}}$ the set of all interpretations on $\mathrm{Prop}$. An interpretation $\mathcal{I} \in \Omega$ is a function $\mathcal{I} : \mathrm{Prop} \rightarrow \{0,1\}$ assigning a truth value $p^{\mathcal{I}}$ to every atomic proposition $p \in \mathrm{Prop}$ and, by extension, a truth value $\phi^{\mathcal{I}}$ to all formulas $\phi \in \mathcal{L}$. We will denote by $[\phi]$ the set of all models of $\phi$, $[\phi] = \{\mathcal{I} : \mathcal{I} \models \phi\}$.

We can give the arguments a structure, and the attack relation is defined in terms of such a structure of the arguments, following the example of [Besnard and Hunter, 2001].

**Definition 7** *An argument is a pair $\langle \Phi, \phi \rangle$, with $\phi \in \mathcal{L}$ and $\Phi \subset \mathcal{L}$, such that*

1. *$\Phi \nvdash \bot$,*
2. *$\Phi \vdash \phi$, and*
3. *$\Phi$ is minimal w.r.t. set inclusion.*

*We say that $\langle \Phi, \phi \rangle$ is an argument for $\phi$. We call $\phi$ the conclusion and $\Phi$ the support of the argument.*

The more specific forms of conflict are called *undercut* and *rebuttal*.

**Definition 8** *(Undercut, Rebuttal) An undercut for an argument $\langle \Phi, \phi \rangle$ is an argument $\langle \Psi, \psi \rangle$ where $\psi = \neg(\phi_1 \wedge \ldots \wedge \phi_n)$ and $\{\phi_1, \ldots, \phi_n\} \subseteq \Phi$. A rebuttal for an argument $\langle \Phi, \phi \rangle$ is an argument $\langle \Psi, \psi \rangle$ iff $\psi \Leftrightarrow \neg\phi$ is a tautology.*

Argument $A_i$ attacks argument $A_j$ where $A_i = \langle \Psi, \psi \rangle$ and $A_j = \langle \Phi, \phi \rangle$ if either $A_1$ undercuts $A_2$ or $A_1$ rebuts $A_2$.

Throughout the paper, we will make the assumption that $\mathcal{A}$ is finite. Indeed, if $\mathcal{A}$ is the set of arguments that has been "received" by an agent, it is very reasonable to assume that the agent, who started operating at some time in the past and has a finite history, may have received, during its finite life, a finite number of arguments from finitely many sources.

Another assumption that we make is that an agent never forgets an argument it has been offered. Therefore, $\mathcal{A}$ may never shrink in time, i.e., if we denote by $\mathcal{A}_t$ the set of arguments received by an agent up to time $t$,

$$t_1 < t_2 \Rightarrow \mathcal{A}_{t_1} \subseteq \mathcal{A}_{t_2}. \quad (1)$$

Given an argument $A \in \mathcal{A}$, we will denote by $\mathrm{src}(A)$ the set of the sources of $A$.

[1] Like in [Benferhat and Kaci, 2003], we adopt the restriction to the finite case in order to use standard definitions of possibilistic logic. Extensions of possibilistic logic to the infinite case are discussed for example in [De Baets *et al.*, 1999].

## 3 Fuzzy Labeling

In order to provide the intuition behind the idea of a fuzzy labeling of the arguments, consider the following dialogue in the context of a murder.

**Example 1** *The judge has to decide whether John has killed Mary. The agents are $Wit_1$ and $Wit_2$, two witnesses, a coroner $Cor$ and the judge $Jud$. Assume the judge completely trusts the two witnesses but he does not quite trust (lower degree of trust) the coroner because it is well-known that he is almost always drunk. The judge starts with argument $A$: "If John did not kill Mary, then John is innocent" where the premise is "If John did not kill Mary" and the conclusion is "John is innocent". Then, the judge listens to the depositions of the two witnesses. $Wit_1$ asserts argument $B$: "I saw John killing Mary, thus John killed Mary". Argument $B$ attacks $A$'s premise so we have an attack $B \rightarrow A$. $Wit_2$ claims $C$: "John was at the theater with me when Mary was killed, thus John did not kill Mary". Argument $C$ attacks $B$'s conclusion and this leads to $C \rightarrow B$. Finally, the judge listens to the deposition of the coroner who asserts $D$: "Mary was killed before 6 p.m., thus when Mary was killed the show was still to begin". Argument $D$ attacks $C$'s premise introducing an attack $D \rightarrow C$. The attack relation is as follows: $D \rightarrow C, C \rightarrow B, B \rightarrow A$.*

Example 1 presents a scenario where the arguments cannot be evaluated from the beginning in the same way because of the degree of trust assigned to their source. In order to account for the fact that arguments may originate from sources that are trusted only to a certain degree, we extend the (crisp) abstract argumentation structure described in Section 2 by allowing gradual membership of arguments in the set of arguments $\mathcal{A}$. In other words, $\mathcal{A}$ is a fuzzy set of trustful arguments, and $\mathcal{A}(A)$, the membership degree of argument $A$ in $\mathcal{A}$, is given by the trust degree of the most reliable (i.e., trusted) source that offers argument $A$[2],

$$\mathcal{A}(A) = \max_{s \in \mathrm{src}(A)} \tau_s, \tag{2}$$

where $\tau_s$ is the degree to which source $s \in \mathrm{src}(A)$ is trusted.

It must be stressed that the fuzzy contribution in our approach is different from the one proposed by [Janssen *et al.*, 2008]. Their fuzzy approach enriches the expressive power of classical argumentation by allowing to represent the relative strength of the attack relations between the arguments, while in our approach the attack relations remains crisp; fuzzyness is introduced to represent uncertainty due to the fact that information sources can also be "just" partially trusted.

[Tang *et al.*, 2010] introduce a framework for decision making where they define trust-extended argumentation graphs in which each premise, inference rule and conclusion is associated to the trustworthiness degree of the source proposing it. Thus, given two arguments rebutting each others, the argument whose conclusion has an higher trust value is accepted. The difference is that in such a framework the "labels", i.e., the trust values, associated to the arguments never change and the arguments are always accepted with the same degree even if they are attacked by more trusted arguments.

This fuzzification of $\mathcal{A}$ provides a natural way of associating strengths to arguments, and suggests rethinking the labeling of an argumentation framework in terms of fuzzy degrees of argument acceptability. [Matt and Toni, 2008] define the strength of the argument the proponent embraces as his long run expected payoff. The difference with our fuzzy labeling is that they compute these strengths from probability distributions on the values of a game. The idea in common with our work is to replace the three-valued labeling with a graded labeling function.

**Definition 9** *(Fuzzy AF-labeling) Let $\langle \mathcal{A}, \rightarrow \rangle$ be an abstract argumentation framework. A fuzzy AF-labeling is a total function $\alpha : \mathcal{A} \rightarrow [0, 1]$.*

Such an $\alpha$ may also be regarded as (the membership function of) the fuzzy set of acceptable arguments: $\alpha(A) = 0$ means the argument is outright unacceptable, $\alpha(A) = 1$ means the argument is fully acceptable, and all cases inbetween are provided for.

Intuitively, the acceptability of an argument should not be greater than the degree to which the arguments attacking it are unacceptable:

$$\alpha(A) \leq 1 - \max_{B:B \rightarrow A} \alpha(B). \tag{3}$$

This is, indeed, a fuzzy reformulation of two basic postulates for reinstatement proposed by [Caminada, 2006] to characterize the labeling of arguments: (1) an argument must be *in* iff all of its attackers are *out*; (2) an argument must be *out* iff there exists an *in* argument that attacks it.

Furthermore, it seems reasonable to require that

$$\alpha(A) \leq \mathcal{A}(A), \tag{4}$$

i.e., an argument cannot be more acceptable than the degree to which its sources are trusted.

By combining the above two postulates, we obtain the following definition.

**Definition 10** *(Fuzzy Reinstatement Labeling) Let $\alpha$ be a fuzzy AF-labeling. We say that $\alpha$ is a fuzzy reinstatement labeling iff, for all arguments $A$,*

$$\alpha(A) = \min\{\mathcal{A}(A), 1 - \max_{B:B \rightarrow A} \alpha(B)\}. \tag{5}$$

We can verify that the fuzzy reinstatement labeling is a generalization of the crisp reinstatement labeling of Definition 5, whose *in* and *out* labels are particular cases corresponding, respectively, to $\alpha(A) = 1$ and $\alpha(A) = 0$. What about the *undec* label in the fuzzy case? One might argue that it corresponds to $\alpha(A) = 0.5$; however, an exam of the case of two arguments attacking each other, $A \rightarrow B$ and $B \rightarrow A$, with $\mathcal{A}(A) = \mathcal{A}(B) = 1$, reveals that any fuzzy reinstatement labeling $\alpha$ must satisfy the equation

$$\alpha(A) = 1 - \alpha(B), \tag{6}$$

which has infinitely many solutions with $\alpha(A) \in [0, 1]$. We can conclude that there are infinitely many degrees of "undecidedness" due to the trustworthiness of the source, of which

[2]Here, we suppose that the agent is optmistic. To represent the behaviour of a pessimistic agent, we should use the min operator, for example.

0.5 is but the most undecided representative. These degrees of "undecidedness" express how much the agent tends to accept those arguments proposed by not fully trusted agents.

Given a fuzzy argumentation framework, how to compute its fuzzy reinstatement labeling? The answer to this question amounts to solving a system of $n$ non-linear equations, where $n = \|\text{supp}(\mathcal{A})\|$, i.e., the number of arguments belonging to some non-zero degree in the fuzzy argumentation framework, of the same form as Equation 5, in $n$ unknown variables, namely, the labels $\alpha(A)$ for all $A \in \text{supp}(\mathcal{A})$. Since iterative methods are usually the only choice for solving systems of non-linear equations, we will resort to this technique, but with an eye to how the labeling is computed in the crisp case. In particular, we draw some inspiration from [Caminada, 2007]'s idea. We start with an all-in labeling (a labeling in which every argument is labeled with the degree it belongs to $\mathcal{A}$). We introduce the notion of illegal labeling for argument $A$ with respect to Definition 10.

**Definition 11** *(Illegal labeling) Let $\alpha$ be a fuzzy labeling and $A$ be an argument. We say that $A$ is illegally labeled iff $\alpha(A) \neq \min\{\mathcal{A}(A), 1 - \max_{B:B\to A} \alpha(B)\}$.*

In order to have an admissible labeling, the absence of illegally labeled arguments is required. As [Caminada, 2007], we need a way of changing the illegal label of an argument, without creating other illegally labeled arguments.

We denote by $\alpha_0 = \mathcal{A}$ the initial labeling, and by $\alpha_t$ the labeling obtained after the $t^{th}$ iteration of the labeling algorithm.

**Definition 12** *Let $\alpha_t$ be a fuzzy labeling. An iteration in $\alpha_t$ is carried out by computing a new labeling $\alpha_{t+1}$ for all arguments $A$ as follows:*

$$\alpha_{t+1}(A) = \frac{1}{2}\alpha_t(A) + \frac{1}{2}\min\{\mathcal{A}(A), 1 - \max_{B:B\to A} \alpha_t(B)\}. \quad (7)$$

Note that Equation 7 guarantees that $\alpha_t(A) \leq \mathcal{A}(A)$ for all arguments $A$ and for each step of the algorithm.

The above definition actually defines a sequence $\{\alpha_t\}_{t=0,1,\ldots}$ of labelings.

**Theorem 1** *The sequence $\{\alpha_t\}_{t=0,1,\ldots}$ defined above converges.*

**Proof 1** *We have to prove that, for all $A$, there exists a real number $L_A \in [0, \mathcal{A}(A)]$ such that, for all $\epsilon > 0$, there exists $N_A$ such that, for every $t > N_A$, $|\alpha_t(A) - L_A| < \epsilon$.*

*The proof is quite straightforward if one assumes the attack relation to be acyclic. In that case, the thesis can be proved by structural induction on the attack relation: the basis is that if argument $A$ is not attacked by any other argument, Equation 7 reduces to*

$$\alpha_{t+1}(A) = \frac{1}{2}\alpha_t(A) + \frac{1}{2}\mathcal{A}(A),$$

*and, since $\alpha_0 = \mathcal{A}(A)$, the sequence is constant and thus trivially converges to $\mathcal{A}(A)$. The inductive step consists of assuming that $\{\alpha_t(B)\}_{t=0,1,\ldots}$ converges for all arguments $B$ such that $B \to A$, and proving that then $\{\alpha_t(A)\}_{t=0,1,\ldots}$ converges as well. If all $\{\alpha_t(B)\}_{t=0,1,\ldots}$ converge, then so does $\{\mu_t(A)\}_t = \{\min\{\mathcal{A}(A), 1 - \max_{B:B\to A} \alpha_t(B)\}\}_t$, i.e., there exists a real number $L_A \in [0, \mathcal{A}(A)]$ such that, for all $\epsilon > 0$, there exists $N_A$ such that, for every $t > N_A$, $|\mu_t(A) - L_A| < \epsilon$, or $L_A - \epsilon < \mu_t(A) < L_A + \epsilon$. Equation 7 reduces to*

$$\alpha_{t+1}(A) = \frac{1}{2}\alpha_t(A) + \frac{1}{2}\mu_t(A),$$

*We have to distinguish two cases. If $\alpha_{t+1}(A) \geq L_A$,*

$$\begin{aligned}
|\alpha_{t+1}(A) - L_A| &= \alpha_{t+1}(A) - L_A = \\
&= \frac{1}{2}\alpha_t(A) + \frac{1}{2}\mu_t(A) - L_A < \\
&< \frac{1}{2}\alpha_t(A) + \frac{1}{2}(L_A + \epsilon) - L_A = \\
&= \frac{1}{2}\alpha_t(A) - \frac{1}{2}L_A + \epsilon/2 = \\
&= \frac{1}{2}(\alpha_t(A) - L_A) + \epsilon/2 \leq \\
&\leq \frac{1}{2}|\alpha_t(A) - L_A| + \epsilon/2.
\end{aligned}$$

*Otherwise, $\alpha_{t+1}(A) < L_A$, and*

$$\begin{aligned}
|\alpha_{t+1}(A) - L_A| &= L_A - \alpha_{t+1}(A) = \\
&= L_A - \frac{1}{2}\alpha_t(A) - \frac{1}{2}\mu_t(A) < \\
&< L_A - \frac{1}{2}\alpha_t(A) - \frac{1}{2}(L_A - \epsilon) = \\
&= \frac{1}{2}L_A - \frac{1}{2}\alpha_t(A) + \epsilon/2 = \\
&= \frac{1}{2}(L_A - \alpha_t(A)) + \epsilon/2 \leq \\
&\leq \frac{1}{2}|\alpha_t(A) - L_A| + \epsilon/2.
\end{aligned}$$

*Therefore, $|\alpha_t(A) - L_A| < |\alpha_0(A) - L_A|2^{-t} + \epsilon 2^{-t} \leq 2^{-t} + \epsilon 2^{-t} = \epsilon_1 + \epsilon_2$.*

*The proof in the general case where attack cycles may exist is based on the idea that convergence in cycles may be proved separately, by assuming that $\{\alpha_t(B)\}_{t=0,1,\ldots}$ converges for all arguments $B$ attacking any of the arguments in the cycle.*

*Let arguments $A_0, A_1, \ldots, A_{n-1}$ form a cycle, i.e., for all $i = 0, \ldots, n-1$, $A_i \to A_{i+1 \mod n}$, and let*

$$u(A_i) = \min\{\mathcal{A}(A_i), 1 - \max_{\substack{B:B\to A_i \\ B\notin\{A_0,\ldots,A_{n-1}\}}} L_B\}$$

*be the upper bound of the feasible values for $\alpha(A_i)$. Note that a cycle with no external arguments attacking arguments of the cycle is a special case, whereby $u(A_i) = \mathcal{A}(A_i)$ for all arguments in the cycle.*

*For every pair of arguments $(A_i, A_{i+1 \mod n})$, for $\alpha$ to be a fuzzy reinstatement labeling it should be*

$$\alpha(A_{i+1 \mod n}) = \min\{u(A_{i+1 \mod n}), 1 - \alpha(A_i)\}$$

*and*

$$\sum_{i=0}^{n-1} \alpha(A_i) \leq \min\left\{\frac{n}{2}, \sum_{i=0}^{n-1} u(A_i)\right\}.$$

*Now, if $\alpha_t$ is not yet a solution of Equation 5, there are two cases:*

| $t$ | $\alpha_t(A)$ | $\alpha_t(B)$ | $\alpha_t(C)$ |
|---|---|---|---|
| 0 | 1 | 0.4 | 0.2 |
| 1 | 0.9 | 0.2 | 0.2 |
| 2 | 0.85 | 0.15 | 0.2 |
| 3 | 0.825 | 0.15 | 0.2 |
| 4 | 0.8125 | 0.1625 | **0.2** |
| 5 | **0.8** | 0.175 | $\downarrow$ |
| 6 | $\downarrow$ | **0.2** | |

Figure 2: Fuzzy labeling on $AF$: $A \to B, B \to C, C \to A$.

1. *either* $\sum_{i=0}^{n-1} \alpha_t(A_i) > \frac{n}{2}$*, and there exists at least an argument* $A_i$ *such that* $\alpha_t(A_i) > 1 - \alpha_t(A_{i+1 \mod n})$*; in this case, then,*

$$\begin{aligned} \min\{u(A_i), 1 - \alpha_t(A_{i+1 \mod n})\} &\leq \\ 1 - \alpha_t(A_{i+1 \mod n}) &< \alpha_t(A_i) \end{aligned}$$

*and* $\alpha_{t+1}(A_i) < \alpha_t(A_i)$*, whence*

$$\sum_{i=0}^{n-1} \alpha_{t+1}(A_i) < \sum_{i=0}^{n-1} \alpha_t(A_i);$$

2. *or* $\sum_{i=0}^{n-1} \alpha_t(A_i) < \frac{n}{2}$*, and there exists at least an argument* $A_i$ *such that*

$$\alpha_t(A_i) < \min\{u(A_i), 1 - \alpha_t(A_{i+1 \mod n})\};$$

*but then* $\alpha_{t+1}(A_i) > \alpha_t(A_i)$*, whence*

$$\sum_{i=0}^{n-1} \alpha_{t+1}(A_i) > \sum_{i=0}^{n-1} \alpha_t(A_i).$$

*Therefore,* $\alpha_t$ *converges for all the arguments in the cycle, and this concludes the proof.*

An example of the calculation of the fuzzy labeling for an odd cycle with three arguments $A$, $B$, and $C$, such that $A \to B, B \to C, C \to A$ and $\mathcal{A}(A) = 1$, $\mathcal{A}(B) = 0.4$, and $\mathcal{A}(C) = 0.2$, is presented in Figure 2.

We may now define the fuzzy labeling of a fuzzy argumentation framework as the limit of $\{\alpha_t\}_{t=0,1,\ldots}$.

**Definition 13** *Let* $\langle \mathcal{A}, \to \rangle$ *be a fuzzy argumentation framework. A fuzzy reinstatement labeling for such argumentation framework is, for all arguments* $A$*,*

$$\alpha(A) = \lim_{t \to \infty} \alpha_t(A). \tag{8}$$

The convergence speed of the labeling algorithm is linear, as the proof of convergence suggests: in practice, a small number of iterations is enough to get so close to the limit that the error is less than the precision with which the membership degrees are represented in the computer.

**Example 2 (Continued)** *Consider again the dialogue in the context of a murder. The judge fully trusts the two witnesses but he assigns a lower degree of trustworthiness to the coroner. The labels of the arguments at the beginning are:* $\alpha(A) = \mathcal{A}(A) = 1$*,* $\alpha(B) = \mathcal{A}(B) = 1$*,* $\alpha(C) = \mathcal{A}(C) = 1$*,* $\alpha(D) = \mathcal{A}(D) = 0.3$*. The fuzzy reinstatement labeling returns the following values:* $\alpha(D) = 0.3$*,* $\alpha(C) = 0.7$*,* $\alpha(B) = 0.3$*, and* $\alpha(A) = 0.7$*.*

# 4 Changing one's mind: rewind

In the proposed framework, belief reinstatement is then guaranteed thanks to the integration of the argumentation framework with the belief-change phase. More precisely, when a new argument arrives, the argumentation framework is updated using the fuzzy labeling algorithm. Therefore, each argument reinstated by the algorithm will induce the reinstatement, to some extent, of the conclusion of the argument in the belief set and of all the formulas that logically follow from the belief set.

The membership function of a fuzzy set describes the more or less possible and mutually exclusive values of one (or more) variable(s). Such a function can then be seen as a possibility distribution [Zadeh, 1978]. If $\pi_x$ is the fuzzy set of possible values of variable $x$, $\pi_x$ is called the possibility distribution associated to $x$; $\pi_x(v)$ is the possibility degree of $x$ being equal to $v$. A possibility distribution for which there exists a completely possible value ($\exists v_0 : \pi(v_0) = 1$) is said to be *normalized*.

**Definition 14** *(Possibility and Necessity Measures) A possibility distribution* $\pi$ *induces a* possibility measure *and its dual* necessity measure*, denoted by* $\Pi$ *and* $N$ *respectively. Both measures apply to a crisp set* $A$ *and are defined as follows:*

$$\Pi(A) = \sup_{s \in A} \pi(s); \tag{9}$$

$$N(A) = 1 - \Pi(\bar{A}) = \inf_{s \in \bar{A}} \{1 - \pi(s)\}. \tag{10}$$

As convincingly argued by [Dubois and Prade, 2009], a *belief* should be regarded as a necessity degree induced by a normalized possibility distribution

$$\pi : \Omega \to [0, 1], \tag{11}$$

which represents a plausibility order of possible states of affairs: $\pi(\mathcal{I})$ is the possibility degree of interpretation $\mathcal{I}$.

Starting from such an insight, a fuzzy reinstatement labeling $\alpha$ determines a set of beliefs in a natural way. Given argument $A = \langle \Phi, \phi \rangle$, let $\mathrm{con}(A)$ denote the conclusion of $A$, i.e., $\mathrm{con}(\langle \Phi, \phi \rangle) = \phi$. The possibility distribution $\pi$ induced by a fuzzy argumentation framework may be constructed by letting, for all interpretation $\mathcal{I}$,

$$\pi(\mathcal{I}) = \min\{1, 1 + \max_{A : \mathcal{I} \models \mathrm{con}(A)} \alpha(A) - \max_{B : \mathcal{I} \not\models \mathrm{con}(B)} \alpha(B)\}. \tag{12}$$

The first maximum in the above equation accounts for the most convincing argument compatible with world $\mathcal{I}$, whereas the second maximum accounts for the most convincing argument against world $\mathcal{I}$. A world will be possible to an extent proportional to the difference between the acceptability of the most convincing argument supporting it and the acceptability of the most convincing argument against it. The world will be considered completely possible if such difference is positive or null, but it will be considered less and less possible (or plausible) as such difference grows more and more negative.

**Theorem 2** *Any* $\pi$ *defined as per Equation 12 is normalized.*

**Proof 2** *Either* $\pi(\mathcal{I}) = 1$ *for all* $\mathcal{I}$*, and* $\pi$ *is trivially normalized, or there exists an interpretation, say* $\mathcal{I}_0$*, such that* $\pi(\mathcal{I}_0) < 1$*. By Equation 12, then, it must be*

$$\max_{A : \mathcal{I}_0 \models \mathrm{con}(A)} \alpha(A) < \max_{B : \mathcal{I}_0 \not\models \mathrm{con}(B)} \alpha(B).$$

*But then, let us consider the complementary interpretation $\overline{\mathcal{I}_0}$, which maps all atoms to a truth value that is the opposite of the truth value they are mapped to by $\mathcal{I}_0$. Clearly, all formulas satisfied by $\mathcal{I}_0$ are not satisfied by $\overline{\mathcal{I}_0}$ and* vice versa*. Therefore,*

$$\pi(\overline{\mathcal{I}_0}) = \min\{1, 1 + \max_{B:\mathcal{I}\not\models \mathrm{con}(B)} \alpha(B) - \max_{A:\mathcal{I}\models \mathrm{con}(A)} \alpha(A)\} = 1.$$

*In other words, if a world is not completely plausible, its opposite must be completely plausible, and for this reason $\pi$ is always normalized.*

### 4.1 Belief Set

The degree to which a given arbitrary formula $\phi \in \mathcal{L}$ is believed can be calculated from the possibility distribution induced by the fuzzy argumentation framework as

$$\mathbf{B}(\phi) = N([\phi]) = 1 - \max_{\mathcal{I}\not\models\phi}\{\pi(\mathcal{I})\}. \tag{13}$$

Such $\mathbf{B}$ may be regarded, at the same time, as a fuzzy modal epistemic operator or as a fuzzy subset of $\mathcal{L}$.

A powerful feature of such an approach based on a possibility distribution is that $\mathbf{B}(\phi)$ can be computed for any formula $\phi$, not just for formulas that are the conclusion of some argument. For instance, if $A$ is an argument whose conclusion is $p$ and $B$ is an argument whose conclusion is $p \supset q$, and $\alpha(A) = \alpha(B) = 1$, then not only $\mathbf{B}(p) = \mathbf{B}(p \supset q) = 1$, but also $\mathbf{B}(q) = 1$, $\mathbf{B}(p \wedge q) = 1$, etc.

Straightforward consequences of the properties of possibility and necessity measures are that $\mathbf{B}(\phi) > 0 \Rightarrow \mathbf{B}(\neg\phi) = 0$, this means that if the agent somehow believes $\phi$ then it cannot believe $\neg\phi$ at all;

$$\mathbf{B}(\top) = 1, \tag{14}$$
$$\mathbf{B}(\bot) = 0, \tag{15}$$
$$\mathbf{B}(\phi \wedge \psi) = \min\{\mathbf{B}(\phi), \mathbf{B}(\psi)\}, \tag{16}$$
$$\mathbf{B}(\phi \vee \psi) \geq \max\{\mathbf{B}(\phi), \mathbf{B}(\psi)\}. \tag{17}$$

### 4.2 Changing Beliefs

We can finally investigate the degree of the agent's belief in terms of the labeling values of the arguments. Let $A$, $B$, $A_0$, and $B_0$ represent arguments, and let $\mu \in (0, 1]$ be a degree of belief. Then, for all $\phi \in \mathcal{L}$,

$$\begin{aligned}
& \mathbf{B}(\phi) \geq \mu \\
\Leftrightarrow\ & \forall \mathcal{I} \not\models \phi \quad \pi(\mathcal{I}) \leq 1 - \mu, \text{ (Eq. 13)} \\
\Leftrightarrow\ & \forall \mathcal{I} \not\models \phi \\
& 1 + \max_{A:\mathcal{I}\models \mathrm{con}(A)} \alpha(A) - \max_{B:\mathcal{I}\not\models \mathrm{con}(B)} \alpha(B) \leq 1 - \mu, \text{ (Eq. 12)} \\
\Leftrightarrow\ & \forall \mathcal{I} \not\models \phi \max_{B:\mathcal{I}\not\models \mathrm{con}(B)} \alpha(B) - \max_{A:\mathcal{I}\models \mathrm{con}(A)} \alpha(A) \geq \mu, \\
\Leftrightarrow\ & \forall \mathcal{I} \not\models \phi\, \exists B_0 : \mathcal{I} \not\models \mathrm{con}(B_0), \forall A : \mathcal{I} \models \mathrm{con}(A), \\
& \alpha(B_0) - \alpha(A) \geq \mu.
\end{aligned}$$

In words, a necessary and sufficient condition for formula $\phi$ to be believed to some extent is that, for all interpretation $\mathcal{I}$ which does not satisfy $\phi$, there exists an argument whose consequence is not satisfied by $\mathcal{I}$ that is more accepted than every argument whose consequence is satisfied by $\mathcal{I}$.

Therefore, the necessary and sufficient condition for formula $\phi$ not to be believed may be stated as follows:

$$\begin{aligned}
& \mathbf{B}(\phi) = 0 \\
\Leftrightarrow\ & \exists \mathcal{I}_0 \not\models \phi, \exists A_0 : \mathcal{I}_0 \models \mathrm{con}(A_0), \forall B : \mathcal{I}_0 \not\models \mathrm{con}(B), \\
& \alpha(A_0) \geq \alpha(B).
\end{aligned}$$

Indeed,

$$\begin{aligned}
& \mathbf{B}(\phi) = 0 \\
\Leftrightarrow\ & \exists \mathcal{I}_0 \not\models \phi : \pi(\mathcal{I}_0) = 1, \\
\Leftrightarrow\ & \exists \mathcal{I}_0 \not\models \phi : \\
& \min\{1, 1 + \max_{A:\mathcal{I}_0\models \mathrm{con}(A)} \alpha(A) - \max_{B:\mathcal{I}_0\not\models \mathrm{con}(B)} \alpha(B)\} = 1, \\
\Leftrightarrow\ & \exists \mathcal{I}_0 \not\models \phi : \max_{A:\mathcal{I}_0\models \mathrm{con}(A)} \alpha(A) \geq \max_{B:\mathcal{I}_0\not\models \mathrm{con}(B)} \alpha(B).
\end{aligned}$$

In this case, a formula $\phi$ is not (or no more) believed by the agent iff there exists an interpretation $\mathcal{I}_0$ which does not satisfy $\phi$ and it is such that there exists an argument whose consequence is satisfied by $\mathcal{I}_0$ and is more accepted than all the arguments whose consequence is not satisfied by $\mathcal{I}_0$.

Therefore, if belief in $\phi$ is lost due to the arrival of an argument $A$ which causes the labeling to change so that $\mathbf{B}(\phi) = 0$, a sufficient condition for reinstatement of $\phi$ is that another argument $A'$ arrives causing the labeling to change so that $\mathbf{B}(\phi) > 0$. However, this does not mean that the previous labeling must be restored, but that it is enough that, for all $\mathcal{I} \not\models \phi$, there exists an argument $B_{\mathcal{I}}$ whose consequence is not satisfied by $\mathcal{I}$, such that $\alpha(B_{\mathcal{I}}) > \alpha(C)$, for all arguments $C$ whose consequence is satisfied by $\mathcal{I}$.

**Example 3 (Continued)** *Suppose the judge finds that $Wit_1$ is little reliable since he was in love with Mary before they broke up because of John. The starting label of argument $B$ becomes $\alpha(B) = \mathcal{A}(B) = 0.2$. The degree to which* $\mathrm{con}(A)$ *is believed at the beginning of the dialogue is* $\mathbf{B}(\mathrm{con}(A)) = 1$*. Then the other three arguments are put forward and the fuzzy reinstatement labeling returns the following values after 53 iterations: $\alpha(D) = 0.3$, $\alpha(C) = 0.7$, $\alpha(B) = 0.2$, and $\alpha(A) = 0.8$. The condition for reinstatement of* $\mathrm{con}(A)$ *is that argument $C$ causes the labeling to change such that* $\mathbf{B}(\mathrm{con}(A)) > 0$*. At the end, the judge believes in John's innocence with a degree given by* $\mathbf{B}(\mathrm{con}(A)) = 0.8$.

## 5 Conclusion

An approach to graded reinstatement in belief revision has been justified and developed. The acceptability of arguments depends on a fuzzy labeling algorithm based on possibility theory. An agent will believe the conclusions of the accepted arguments, as well as their consequences. Arguments reinstatement induces the reinstatement of the beliefs grounded on these arguments. Arguments and beliefs are reinstated depending on the trustworthiness of the sources proposing them. The framework can be further improved following two directions: (i) specifying the trustworthiness degree by a cognitive model of trust and, (ii) embedding the proposed framework into an integrated one where also desires and, afterwards, goals are taken into account. This is left for future work.

[Rahwan *et al.*, 2010] have recently presented a cognitive approach to the analysis of the reinstatement notion. Given a number of psychological experiments, they show that reinstatement does not yield the full expected recovery of the attacked argument. Our theory might explain their results by assuming that the fuzzy reinstatement labeling can assign to the attacked argument a value which does not yield the full recovery of its starting value. However, according to our proposal, the new value of the attacked argument may also be higher than the original, if the source proposing the argument which reinstates the attacked argument is more trusted than the source of the attacked argument.

## References


[Amgoud and Prade, 2004] L. Amgoud and H. Prade. Reaching agreement through argumentation: A possibilistic approach. In Didier Dubois, Christopher A. Welty, and Mary-Anne Williams, editors, *KR*, pages 175–182. AAAI Press, 2004.

[Benferhat and Kaci, 2003] S. Benferhat and S. Kaci. Logical representation and fusion of prioritized information based on guaranteed possibility measures: application to the distance-based merging of classical bases. *Artif. Intell.*, 148(1-2):291–333, 2003.

[Besnard and Hunter, 2001] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artif. Intell.*, 128(1-2):203–235, 2001.

[Caminada, 2006] M. Caminada. On the issue of reinstatement in argumentation. In *Procs. of JELIA*, volume 4160 of *LNCS*, pages 111–123. Springer, 2006.

[Caminada, 2007] M. Caminada. An algorithm for computing semi-stable semantics. In *Procs. of ECSQARU*, volume 4724 of *LNCS*, pages 222–234. Springer, 2007.

[Cayrol *et al.*, 2010] C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *J. Artif. Intell. Res. (JAIR)*, 38:49–84, 2010.

[da Costa Pereira and Tettamanzi, 2010] C. da Costa Pereira and A. Tettamanzi. An integrated possibilistic framework for goal generation in cognitive agents. In *Procs. of AAMAS*, pages 1239–1246, 2010.

[De Baets *et al.*, 1999] B. De Baets, E. Tsiporkova, and R. Mesiar. Conditioning in possibility theory with strict order norms. *Fuzzy Sets Syst.*, 106(2):221–229, 1999.

[Delgrande *et al.*, 2006] J. Delgrande, D. Dubois, and J. Lang. Iterated revision as prioritized merging. In *Procs. of KR*, pages 210–220, 2006.

[Dragoni and Giorgini, 1996] A. Dragoni and P. Giorgini. Belief revision through the belief-function formalism in a multi-agent environment. In *Procs. of ATAL*, volume 1193 of *LNCS*, pages 103–115. Springer, 1996.

[Dubois and Prade, 1997] D. Dubois and H. Prade. A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. *Int. J. Approx. Reasoning*, 17(2-3):295–324, 1997.

[Dubois and Prade, 2009] D. Dubois and H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets Syst.*, 160(10):1355–1366, 2009.

[Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[Falappa *et al.*, 2009] M. Falappa, G. Kern-Isberner, and G. Simari. Belief revision and argumentation theory. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, pages 341–360, 2009.

[Gabbay *et al.*, 2003] D. Gabbay, G. Pigozzi, and J. Woods. Controlled revision - an algorithmic approach for belief revision. *J. Log. Comput.*, 13(1):3–22, 2003.

[Jakobovits and Vermeir, 1999] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *J. Log. Comput.*, 9(2):215–261, 1999.

[Janssen *et al.*, 2008] J. Janssen, M. De Cock, and D. Vermeir. Fuzzy argumentation frameworks. In *Procedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2008)*, pages 513–520, 2008.

[Matt and Toni, 2008] P.-A. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In Steffen Hölldobler, Carsten Lutz, and Heinrich Wansing, editors, *JELIA*, volume 5293 of *LNCS*, pages 285–297. Springer, 2008.

[Paglieri and Castelfranchi, 2006] F. Paglieri and C. Castelfranchi. *Arguing on the Toulmin model*, chapter The Toulmin Test: Framing argumentation within belief revision theories, pages 359–377. Berlin, Springer, 2006.

[Quignard and Baker, 1997] M. Quignard and M. Baker. Modelling argumentation and belief revision in agent interactions. In *Procs. of ECCS*, pages 85–90, 1997.

[Rahwan *et al.*, 2010] I. Rahwan, M. Madakkatel, J.-F. Bonnefon, R. Awan, and S. Abdallah. Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. *Cognitive Science*, 34(8):1483–1502, 2010.

[Tang *et al.*, 2010] Y. Tang, K. Cai, E. Sklar, P. McBurney, and S. Parsons. A system of argumentation for reasoning about trust. In *Procs. of EUMAS*, 2010.

[Verheij, 2003] B. Verheij. Artificial argument assistants for defeasible argumentation. *Artif. Intell.*, 150(1-2):291–324, 2003.

[Zadeh, 1978] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.