Making Others Believe What They Want

Guido Boella, Célia da Costa Pereira, Andrea G. B. Tettamanzi, and Leendert van der Torre

Abstract We study the interplay between argumentation and belief revision within the MAS framework. When an agent uses an argument to persuade another one, he must consider not only the proposition supported by the argument, but also the overall impact of the argument on the beliefs of the addressee. Different arguments lead to different belief revisions by the addressee. We propose an approach whereby the best argument is defined as the one which is both rational and the most appealing to the addressee.

1 A motivating example

Galbraith [5] put forward examples of public communication where speakers have to address a politically oriented audience. He noticed how it is difficult to propose them views which contrast with their goals, values, and what they already know.

Speaker $\mathscr{S}$, a financial advisor, has to persuade addressee $\mathscr{R}$, an investor, who desires to invest a certain amount of money (*im*). $\mathscr{S}$ has two alternative arguments in support of a proposition *wd* ("The dollar is weak") he wants $\mathscr{R}$ to believe, one based on $bt \rightarrow wd$ and one on $hb \rightarrow wd$:

1. "The dollar is weak (*wd*) since the balance of trade is negative (*bt*), due to high import (*hi*)" ($a = \langle \{bt \rightarrow wd, hi \rightarrow bt, hi\} \rangle$)

Guido Boella
Universita di Torino, Dipartimento di Informatica 10149, Torino, Cso Svizzera 185, Italy, e-mail: guido@di.unito.it

Célia da Costa Pereira and Andrea G. B. Tettamanzi
Università degli Studi di Milano, Dip. Tecnologie dell'Informazione, Via Bramante 65, I-26013 Crema (CR), Italy, e-mail: {pereira,tettamanzi}@dti.unimi.it

Leendert van der Torre
Université du Luxembourg, Computer Science and Communication L-1359 Luxembourg, rue Richard Coudenhove – Kalergi 6, Luxembourg, e-mail: leon.vandertorre@uni.lu

2. "The dollar is weak (*wd*) due to the housing bubble (*hb*) created by excess subprime mortgages (*sm*)" ($hb \rightarrow wd$, $sm \rightarrow hb$, *sm*). And to the reply of $\mathscr{R}$: "There is no excess of subprime mortgages (*sm*) since the banks are responsible (*rb*)" ($rb \rightarrow \neg sm$, *rb*), $\mathscr{S}$ counters that "The banks are not responsible (*rb*) as the Enron case shows (*ec*)" ($ec \rightarrow \neg rb$, *ec*).

Assume that both agents consider a supported proposition stronger than an unsupported one (e.g., $ec \rightarrow \neg rb$ prevails on *rb* alone). Although, from a logical point of view, both arguments make the case for *wd*, they are very different if we consider other dimensions concerning the addressee $\mathscr{R}$. For example, even if $\mathscr{R}$ could accept *wd*, other parts of the arguments have different impacts.

Accepting the arguments implies not only believing *wd*, but also the whole argument from which *wd* follows (unless we have an irrational agent which accepts the conclusion of an argument but not the reasons supporting the conclusion). This means that $\mathscr{R}$ undergoes a phase of belief revision to accept the support of the argument, resulting in a new view of the world. Before dropping his previous view of the world and adopting the new one, he has to compare them.

- The state of the world resulting from the revision is less promising from the point of view of the possibility for $\mathscr{R}$ of reaching his goals. E.g., if the banks are not responsible, it is difficult to achieve his goal of investing money *im*.
- The state of the world resulting from the revision contrasts with his values. E.g., he has a subprime mortgage and he does not like a world where subprime mortgages are risky due to their excess.
- He never heard about $hb \rightarrow wd$, even if he trusts $\mathscr{S}$; this is new information for him.

Thus $\mathscr{R}$ is probably leaning to accept the first argument which does not interact with his previous goals and beliefs, rather than to accept the second one, which, above all, depicts a scenario which is less promising for his hopes of making money by investing. Thus, a smart advisor, which is able to figure out the profile of the investor, will resort to the first argument rather than to the second one.

Even if such evaluation of $\mathscr{R}$'s in deciding what to believe can lead to partially irrational decisions, this is what happens in humans. Both economists like Galbraith and cognitive scientists like Castelfranchi [8] support this view. Thus, $\mathscr{S}$ should take advantage of this mechanism of reasoning.

In particular, an agent could pretend to have accepted the argument at the public level, since he cannot reply anymore to the persuader and he does not want to appear irrational. However, privately, and in particular when the time comes to make a decision, he will stick to his previous beliefs.

For this reason, if we want to build agents which are able to interact with humans, or believable agents, or if we want to use agent models as formal models for phenomena which are studied informally in other fields like economics, sociology, and cognitive science, and, moreover, to avoid that our agents are cheated by other agents which exploit mechanisms like the one proposed here, these phenomena must be studied.

2 Argumentation Theory

We adopt a simple framework for argumentation along the lines of Dung's original proposal [4] by instantiating the notion of argument as an explanation-based argument. Given a set of formulas L, an argument over L is a pair $A = \langle H, h \rangle$ such that $H \subseteq L$, H is consistent, $H \vdash h$, and H is minimal (for set inclusion) among the sets satisfying the former three conditions. On the set of arguments $\mathtt{Arg}$, a priority relation $\succeq$ is defined, $A_1 \succeq A_2$ meaning that A_1 has priority over A_2.

Let $A_1 = \langle H_1, h_1 \rangle$ and $A_2 = \langle H_2, h_2 \rangle$ be two arguments. A_1 undercuts A_2, in symbols $A_1 \rightsquigarrow A_2$, if $\exists h'_2 \in H_2$ such that $h_1 \equiv \neg h'_2$. A_1 rebuts A_2, in symbols $A_1 \multimap A_2$, if $h_1 \equiv \neg h_2$ (note that $\multimap$ is symmetric); finally, A_1 attacks A_2, in symbols $A_1 \rightleftharpoons A_2$, if (i) $A_1 \multimap A_2$ or $A_1 \rightsquigarrow A_2$ and, (ii) if $A_2 \multimap A_1$ or $A_2 \rightsquigarrow A_1$, $A_2 \not\succ A_1$.

The semantics of Dung's argumentation framework is based on the two notions of defence and conflict-freeness.

Definition 1. A set of arguments S defends an argument A iff, for each argument $B \in \mathtt{Arg}$ such that $B \rightleftharpoons A$, there exists an argument $C \in S$ such that $C \rightleftharpoons B$.

Definition 2. A set of arguments S is conflict-free iff there are no $A, B \in S$ such that $A \rightleftharpoons B$.

The following definition summarizes various semantics of acceptable arguments proposed in the literature. The output of the argumentation framework is derived from the set of acceptable arguments which are selected with respect to an acceptability semantics.

Definition 3. Let $S \subseteq \mathtt{Arg}$.

- S is *admissible* iff it is conflict-free and defends all its elements.
- A conflict-free S is a *complete extension* iff $S = \{A \mid S \text{ defends } A\}$.
- S is a *grounded extension* iff it is the smallest (for set inclusion) complete extension.
- S is a *preferred extension* iff it is a maximal (for set inclusion) complete extension.
- S is a *stable extension* iff it is a preferred extension that attacks all arguments in $\mathtt{Arg} \backslash S$.

In this paper we use the unique grounded extension, written as $E(\mathtt{Arg}, \succeq)$. Many properties and relations among these semantics have been studied by Dung and others.

Example 1. The example of Section 1 can be formalized as follows in terms of arguments.

$a = \langle \{bt \rightarrow wd, hi \rightarrow bt, hi\}, wd \rangle$, $b = \langle \{eg \rightarrow \neg hi, eg\}, \neg hi) \rangle$,
$c = \langle \{de \rightarrow \neg eg, de\}, \neg eg \rangle$, $c \succeq b$, $d = \langle \{hb \rightarrow wd, sm \rightarrow hb, sm\}, wd \rangle$,
$e = \langle \{rb \rightarrow \neg sm, rb\}, \neg sm \rangle$, $f = \langle \{ec \rightarrow \neg rb, ec\}, \neg rb \rangle$, $f \succeq e$,
$b \rightsquigarrow a$, $c \rightsquigarrow b$, $e \rightsquigarrow d$, $f \rightsquigarrow e$,
$\mathtt{Arg} = \{a, b, c, d, e, f\}$,
$E(\mathtt{Arg}, \succeq) = \{a, c, d, f\}$.

3 Arguments and Belief Revision

Belief revision is the process of changing beliefs to take into account a new piece of information. Traditionally the beliefs are modelled as propositions and the new piece of information is a proposition. In our model, instead, the belief base is made of arguments, and the new information is an argument too.

Let $*$ be an argumentative belief revision operator, it is defined as the addition of the new argument to the base as the one with the highest priority. Given $A = \langle H, h\rangle$, a base of arguments Q and a priority relation $\succeq_Q$ over Q:

$$\langle Q, \succeq_Q\rangle * A = \langle Q \cup \{A\}, \succeq_{(Q,\{A\})}\rangle \tag{1}$$

where $\succeq_Q \subset \succeq_{(Q,\{A\})} \wedge \forall A' \in Q\, A \succ_{(Q,\{A\})} A'$.

The new belief set can be derived from the new extension $E(Q \cup \{A\}, \succeq_{(Q,\{A\})})$ as the set of conclusions of arguments:

$$B(Q \cup \{A\}, \succeq_{(Q,\{A\})}) = \{h \mid \exists \langle H, h\rangle \in E(Q \cup \{A\}, \succeq_{(Q,\{A\})})\}. \tag{2}$$

Note that, given this definition, there is no warranty that the conclusion h of argument A is in the belief set; indeed, even if A is now the argument with highest priority, in the argument set Q there could be some argument A' such that $A' \rightleftharpoons A$. An argument $A' = \langle H', h'\rangle \multimap A$ (i.e., $h' \equiv \neg h$) would not be able to attack A, since $A \succ_Q A'$ by definition of revision. Instead, if $A' \rightsquigarrow A$, it is possible that A does not undercut or rebut A' in turn, and, thus, $A' \rightleftharpoons A$, possibly putting it outside the extension if no argument defends it against A'.

Success can be ensured only if the argument A can be supported by a set of arguments S with $\succeq_S$ which, once added to Q, can defend A in Q and defend themselves too.

Thus, it is necessary to extend the definition above to sets of arguments, to allow an argument to be defended:

$$\langle Q, \succeq_Q\rangle * \langle S, \succeq_S\rangle = \langle Q \cup S, \succeq_{(Q,S)}\rangle \tag{3}$$

where the relative priority among the arguments in S is preserved, and they have priority over the arguments in Q:

$\succeq_Q \subset \succeq_{(Q,S)} \wedge$
$\forall A', A'' \in S\, A' \succ_{(Q,S)} A'' \mathit{iff} A' \succ_S A'' \wedge$
$\forall A \in S, \forall A' \in Q\, A \succ_{(Q,S)} A'$.

Example 2. $Q = \{e\}, S = \{d, f\}, d \succ_S f, f \succ_S d,$
$\langle Q, \succeq_Q\rangle * S = \langle Q \cup S, \succeq_{(Q,S)}\rangle,$
$E(Q \cup S, \succeq_{(Q,S)}) = \{d, f\},$
$d \succ_{(Q,S)} e, f \succ_{(Q,S)} e, d \succ_{(Q,S)} f, f \succ_{(Q,S)} d,$
$B(E(\{d, e, f\}, \succeq_{(Q,S)})) = \{wd, sm\}.$

4 An Abstract Agent Model

The basic components of our language are *beliefs* and *desires*. Beliefs are represented by means of an *argument base*. A belief set is a finite and consistent set of propositional formulas describing the information the agent has about the world and internal information. Desires are represented by means of a *desire set*. A desire set consists of a set of propositional formulas which represent the situations the agent would like to achieve. However, unlike the belief set, a desire set may be inconsistent, e.g., $\{p, \neg p\}$.

Let $\mathscr{L}$ be a propositional language.

Definition 4. The agent's desire set is a possibly inconsistent finite set of sentences denoted by D, with $D \subseteq \mathscr{L}$.

Goals, in contrast to desires, are represented by consistent desire sets.

We assume that an agent is equipped with two components:

- an argument base $\langle \mathtt{Arg}, \succeq_{\mathtt{Arg}} \rangle$ where $\mathtt{Arg}$ is a set of arguments and $\succeq_{\mathtt{Arg}}$ is a priority ordering on arguments.
- a desire set: $D \subseteq \mathscr{L}$;

The mental state of an agent is described by a pair $\Sigma = \langle \langle \mathtt{Arg}, \succeq_{\mathtt{Arg}} \rangle, D \rangle$. In addition, we assume that each agent is provided with a goal selection function G, and a belief revision operator $*$, as discussed below.

Definition 5. We define the belief set, B, of an agent, i.e., the set of all propositions in $\mathscr{L}$ the agent believes, in terms of the extension of its argument base $\langle \mathtt{Arg}, \succeq_{\mathtt{Arg}} \rangle$:

$$B = B(\mathtt{Arg}, \succeq_{\mathtt{Arg}}) = \{h \mid \exists \langle H, h \rangle \in E(\mathtt{Arg}, \succeq_{\mathtt{Arg}})\}.$$

We will denote by $\Sigma_{\mathscr{S}}$, $\mathtt{Arg}_{\mathscr{S}}$, $E(\mathtt{Arg}_{\mathscr{S}}, \succeq_{\mathtt{Arg}})$ and $B_{\mathscr{S}}$, respectively, the mental state, the argument base, the extension of $\mathtt{Arg}_S$, and the belief set of an agent $\mathscr{S}$.

In general, given a problem, not all goals are *achievable*, i.e. it is not always possible to construct a plan for each goal. The goals which are not achievable or those which are not chosen to be achieved are called *violated goals*. Hence, we assume a problem-dependent function $\mathscr{V}$ that, given a belief base B and a goal set $D' \subseteq D$, returns a set of couples $\langle D^a, D^v \rangle$, where D^a is a maximal subset of achievable goals and D^v is the subset of violated goals and is such that $D^v = D' \setminus D^a$. Intuitively, by considering violated goals we can take into account, when comparing candidate goal sets, what we lose from not achieving goals.

In order to act an agent has to take a decision among the different sets of goals he can achieve.

The aim of this section is to illustrate a qualitative method for goal comparison in the agent theory. More precisely, we define a qualitative way in which an agent can choose among different sets of candidate goals. Indeed, from a desire set D, several candidate goal sets D_i, $1 \leq i \leq n$, may be derived. How can an agent choose

among all the possible D_i? It is unrealistic to assume that all goals have the same priority. We use the notion of preference (or urgence) of desires to represent how relevant each goal should be for the agent depending, for instance, on the reward for achieving it. The idea is that an agent should choose a set of candidate goals which contains the greatest number of achievable goals (or the least number of violated goals).

We assume we dispose of a total pre-order $\succeq$ over an agent's desires, where $\phi \succeq \psi$ means desire ϕ is at least as preferred as desire ψ.

The $\succeq$ relation can be extended from goals to sets of goals. We have that a goal set D_1 is preferred to another one D_2 if, considering only the goals occurring in either set, the most preferred goals are in D_1. Note that $\succeq$ is connected and therefore a total pre-order, i.e., we always have $D_1 \succeq D_2$ or $D_2 \succeq D_1$ (or both).

Definition 6. Goal set D_1 is at least as important as goal set D_2, denoted $D_1 \succeq D_2$ iff the list of desires in D_1 sorted by decreasing preference is lexicographically greater than the list of desires in D_2 sorted by decreasing importance. If $D_1 \succeq D_2$ and $D_2 \succeq D_1$, D_1 and D_2 are said to be indifferent, denoted $D_1 \sim D_2$.

However, we also need to be able to compare the mutual exclusive subsets (achievable and violated goals) of the considered candidate goal, as defined below.

We propose two methods to compare couples of goal sets.

Given the $\succeq_D$ criterion, a couple of goal sets $\langle D_1^a, D_1^v \rangle$ is at least as preferred as the couple $\langle D_2^a, D_2^v \rangle$, noted $\langle D_1^a, D_1^v \rangle \succeq_D \langle D_2^a, D_2^v \rangle$ iff $D_1^a \succeq D_2^a$ and $D_1^v \preceq D_2^v$.
$\succeq_D$ is reflexive and transitive but partial. $\langle D_1^a, D_1^v \rangle$ is strictly preferred to $\langle D_2^a, D_2^v \rangle$ in two cases:

1. $D_1^a \succeq D_2^a$ and $D_1^v \prec D_2^v$, or
2. $D_1^a \succ D_2^a$ and $D_1^v \preceq D_2^v$.

They are indifferent when $D_1^a = D_2^a$ and $D_1^v = D_2^v$. In all the other cases, they are not comparable.

Given the $\succeq_{Lex}$ criterion, a couple of goal sets $\langle D_1^a, D_1^v \rangle$ is at least as preferred as the couple $\langle D_2^a, D_2^v \rangle$ (noted $\langle D_1^a, D_1^v \rangle \succeq_{Lex} \langle D_2^a, D_2^v \rangle$) iff $D_1^a \sim D_2^a$ and $D_1^v \sim D_2^v$; or there exists a $\phi \in \mathscr{L}$ such that both the following conditions hold:

1. $\forall \phi' \succeq \phi$, the two couples are indifferent, i.e., one of the following possibilities holds: (a) $\phi' \in D_1^a \cap D_2^a$; (b) $\phi' \notin D_1^a \cup D_1^v$ and $\phi' \notin D_2^a \cup D_2^v$; (c) $\phi' \in D_1^v \cap D_2^v$.
2. Either $\phi \in D_1^a \setminus D_2^a$ or $\phi \in D_2^v \setminus D_1^v$.

$\succeq_{Lex}$ is reflexive, transitive, and total.

In general, given a set of desires D, there may be many possible candidate goal sets. An agent in state $\Sigma = \langle \mathtt{Arg}, D \rangle$ must select precisely one of the most preferred couples of achievable and violated goals.

Let us call G the function which maps a state Σ into the couple $\langle D^a, D^v \rangle$ of goal sets selected by an agent in state Σ. G is such that, $\forall \Sigma$, if $\langle \bar{D}^a, \bar{D}^v \rangle$ is a couple of goal sets, then $G(\Sigma) \succeq \langle \bar{D}^a, \bar{D}^v \rangle$, i.e., a rational agent always selects one of the most preferable couple of candidate goal sets [3].

5 An Abstract Model of Speaker-Receiver Interaction

Using the above agent model, we consider two agents, $\mathscr{S}$, the speaker, and $\mathscr{R}$, the receiver. $\mathscr{S}$ wants to convince $\mathscr{R}$ of some proposition p.

How does agent $\mathscr{S}$ construct a set of arguments S? Of course, $\mathscr{S}$ could include all the arguments in its base, but in this case it would risk to make his argumentation less appealing and thus to make $\mathscr{R}$ refuse to revise its beliefs, as discussed in the next section. Thus, we require that the set of arguments S to be communicated to $\mathscr{R}$ is minimal: even if there are alternative arguments for p, only one is included.

We include the requirement that S is chosen using arguments which are not already believed by $\mathscr{R}$. S is a minimal set among the T defined in the following way:

$$T \subseteq \mathtt{Arg}_{\mathscr{S}} \wedge B(\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}} \rangle * \langle T, \succeq_{\mathscr{S}} \rangle)) \vdash p. \tag{4}$$

Example 3. $S = \{a,c\}, p = wd, \mathtt{Arg}_{\mathscr{R}} = \{b\} E(\mathtt{Arg}_{\mathscr{R}} \cup S, \succeq_{(\mathtt{Arg}_{\mathscr{R}},S)}) = \{a,c\}$, $B(E(\mathtt{Arg}_{\mathscr{R}} \cup S, \succeq_{(\mathtt{Arg}_{\mathscr{R}},S)})) = \{wd, \neg eg\}$.

This definition has two shortcomings: first, such an S may not exist, since T could be empty. There is no reasonable way of assuring that $\mathscr{S}$ can always convince $\mathscr{R}$: as we discussed in Section 3, no success can be assumed.

Second, in some cases arguments in $E(\mathtt{Arg}_{\mathscr{R}} \cup S, \succeq_{(\mathtt{Arg}_{\mathscr{R}},S)})$ may be among the ones believed by $\mathscr{R}$ but not by $\mathscr{S}$. If they contribute to prove p, there would be a problem: $\exists A \in \mathtt{Arg}_{\mathscr{R}} \backslash \mathtt{Arg}_{\mathscr{S}}\ B(E((\mathtt{Arg}_{\mathscr{R}} \backslash \{A\}) \cup S, \succeq_{(\mathtt{Arg}_{\mathscr{R}},S)})) \nvdash p$

This would qualify $\mathscr{S}$ as a not entirely sincere agent, since he would rely (even if he does not communicate them explicitly) on some arguments he does not believe, which are used in the construction of the extension from which p is proved.

The second problem, instead, can be solved in the following way, by restricting set S not to require arguments not believed by $\mathscr{S}$ to defend S. S is now a minimal T such that $T \subseteq \mathtt{Arg}_{\mathscr{S}}$ and $B(\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}} \rangle * \langle T, \succeq_{\mathscr{S}} \rangle) \vdash p$ and $\neg \exists A \in \mathtt{Arg}_{\mathscr{R}} \backslash \mathtt{Arg}_{\mathscr{S}}\ B(\langle \mathtt{Arg}_{\mathscr{R}} \backslash \{A\}, \succeq_{\mathtt{Arg}_{\mathscr{R}}} \rangle * \langle T, \succeq_{\mathscr{S}} \rangle) \nvdash p$

Example 4. $\mathtt{Arg}_{\mathscr{S}} = \{a,c,i\}, \mathtt{Arg}_{\mathscr{R}} = \{b,g,h\}, g \rightsquigarrow c, h \rightleftharpoons g, i \rightleftharpoons g.$
If $S = \{a,c\}, p = wd$: $E(\mathtt{Arg}_{\mathscr{R}} \cup S, \succeq_{(\mathtt{Arg}_{\mathscr{R}},S)}) = \{a,c,h\}$,
$B(\{a,b,c,g,h\}) = \{wd, \neg eg \ldots\}$.
If $S = \{a,c,i\}, p = wd$: $E(\mathtt{Arg}_{\mathscr{R}} \cup S, \succeq_{(\mathtt{Arg}_{\mathscr{R}},S)}) = \{a,c,i\}$,
$B(\{a,b,c,g,h,i\}) = \{wd, \neg eg \ldots\}$.

The belief revision system based on argumentation (see Section 2), is used to revise the public face of agents: the agents want to appear rational (otherwise they lose their status, reliability, trust, etc.) and, thus, when facing an acceptable argument (i.e., they do not know what to reply) have to admit that they believe it and to revise the beliefs which are inconsistent with it.

We want to model social interactions among agent which do not necessarily tell the truth or trust each other completely, although they may pretend to. In such a setting, an agent revises its private beliefs only if someone provides an acceptable argument in the sense of Section 2.

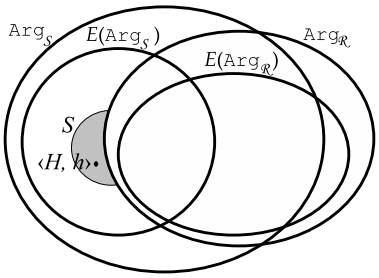

Fig. 1 A diagram of mutual inclusion relations among the belief bases and sets involved in the interaction between $\mathscr{S}$ and $\mathscr{R}$.

Thus, while publicly an agent must pretend to be rational and thus shall revise its public belief base according to the system discussed in Section 3, nothing forbids an agent to privately follow other types of rules, not even necessarily rational. As a worst-case scenario (from $\mathscr{S}$'s standpoint), we assume that $\mathscr{R}$ uses a belief revision system based on Galbraith's notion of conventional wisdom discussed in [2] as a proposal to model the way an irrational (but realistic) agent might revise its private beliefs.

The idea is that different sets of arguments $S_1, \ldots, S_n$ lead to different belief revisions $\langle \mathtt{Arg}, \succeq_{\mathtt{Arg}} \rangle * \langle S_1, \succeq_{S_1} \rangle, \ldots, \langle \mathtt{Arg}_{\mathtt{Arg}}, \succeq_{\mathtt{Arg}} \rangle * \langle S_n, \succeq_{S_n} \rangle$. $\mathscr{R}$ will privately accept the most appealing argument, i.e., the S_i which maximizes the preferences according to the notion of Galbraith's conventional wisdom.

In order to formalize this idea, we have to define an order of *appeal* on sets of beliefs.

Definition 7. Let $\mathtt{Arg}_1$ and $\mathtt{Arg}_2$ be two argument bases. $\mathtt{Arg}_1$ is *more appealing* than $\mathtt{Arg}_2$ to an agent, with respect to the agent's desire set D, in symbols $\mathtt{Arg}_1 \succeq \mathtt{Arg}_2$, if and only if $G(\langle \langle \mathtt{Arg}_1, \succeq_{\mathtt{Arg}_1} \rangle, D \rangle) \succeq G(\langle \langle \mathtt{Arg}_2, \succeq_{\mathtt{Arg}_2} \rangle, D \rangle)$.

We will denote by $\bullet$ the private, CW-based belief revision operator. Given an acceptable argument set S,

$$\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}} \rangle \bullet \langle S, \succeq_S \rangle \in \{ \langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}} \rangle, \langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}} \rangle * \langle S, \succeq_S \rangle \}.$$

This definition is inspired to indeterministic belief revision [6]: "Most models of belief change are deterministic. Clearly, this is not a realistic feature, but it makes the models much simpler and easier to handle, not least from a computational point of view. In indeterministic belief change, the subjection of a specified belief base to a specified input has more than one admissible outcome.

Indeterministic operators can be constructed as sets of deterministic operations. Hence, given n deterministic revision operators $*_1, *_2, \ldots, *_n$, $* = \{*_1, *_2, \ldots, *_n\}$ can be used as an indeterministic operator."

We then define the notion of appealing argument, i.e., an argument which is preferred by the receiver $\mathscr{R}$ to the current state of its beliefs:

Definition 8. Let S be a minimal set of arguments that supports $A = \langle H, p\rangle$, such that S defends A and defends itself, as defined in the previous section:
$\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle \bullet \langle S, \succeq_S\rangle = \langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S, \succeq_S\rangle$,
i.e., $\mathscr{R}$ privately accepts revision $\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S, \succeq_S\rangle$, if

$$\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S, \succeq_S\rangle \succeq \langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle$$

otherwise $\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle \bullet \langle S, \succeq_S\rangle = \langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle$.

Example 5. The investor of our example desires investing money. Assuming this is his only desire, we have $D_{\mathscr{R}} = \{im\}$. Now, the advisor $\mathscr{S}$ has two sets of arguments to persuade $\mathscr{R}$ that the dollar is weak, namely $S_1 = \{a, c\}$ and $S_2 = \{d, f\}$. Let us assume that, according to the "planning module" of $\mathscr{R}$,

$$\mathscr{V}(\langle\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S_1, \succeq_{S_1}\rangle, D_{\mathscr{R}}\rangle) = \langle\{im\}, \emptyset\rangle,$$
$$\mathscr{V}(\langle\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S_2, \succeq_{S_2}\rangle, D_{\mathscr{R}}\rangle) = \langle\emptyset, \{im\}\rangle.$$

Therefore, $G(\langle\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S_1, \succeq_{S_1}\rangle, D_{\mathscr{R}}\rangle) \succeq G(\langle\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle * \langle S_2, \succeq_{S_2}\rangle, D_{\mathscr{R}}\rangle)$, because, by revising with $S_1 = \{a, c\}$, $\mathscr{R}$'s desire *im* is achievable.

A necessary and sufficient condition for the public and private revisions to coincide is thus that the set of arguments S used to persuade an agent is the most appealing for the addressee, if one exists.

Since CW-based belief revision is indeterministic and not revising is an alternative, $\mathscr{R}$ decides whether to keep the *status quo* of his beliefs or to adopt the belief revision resulting from the arguments proposed by $\mathscr{S}$.

Seen from $\mathscr{S}$'s standpoint, the task of persuading $\mathscr{R}$ of p is about comparing $\mathscr{R}$'s belief revisions resulting from the different sets of arguments supporting p and acceptable by $\mathscr{R}$, and choosing the set of arguments that appeals most to $\mathscr{R}$.

To define the notion of the most appealing set of arguments, we need to extend the order of appeal $\succeq$ to sets of arguments.

Definition 9. Let S_1 and S_2 be two sets of arguments that defend themselves; S_1 is more appealing to $\mathscr{R}$ than S_2, in symbols $S_1 \succeq_{\mathscr{R}} S_2$, if and only if

$$\langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle \bullet \langle S_1, \succeq_{S_1}\rangle \succeq \langle \mathtt{Arg}_{\mathscr{R}}, \succeq_{\mathtt{Arg}_{\mathscr{R}}}\rangle \bullet \langle S_2, \succeq_{S_2}\rangle.$$

The most appealing set of arguments S_p^* for persuading $\mathscr{R}$ of p, according to conventional wisdom, is, among all minimal sets of arguments S that support an $A = \langle H, p\rangle$, such that S defends A and S defends itself as defined in Section 5, the one that is maximal with respect to the appeal $\succeq_{\mathscr{R}}$, i.e., such that $S_p^* \succeq_{\mathscr{R}} S$.

6 Conclusions

We studied how to choose arguments in persuasion to maximize their acceptability with respect to their receiver. In some applications, when agents have to interact with

human users who act in a non fully rational way, like, e.g., following the principle of conventional wisdom, it is necessary to model such a behavior.

To model the process of selecting acceptable arguments, in this paper:

- We derive the beliefs of an agent from a base of arguments. An agent believes the propositions which are supported by the arguments of the grounded extension of its argument base.
- We propose a definition of belief revision of an argument base as an expansion of the base with the new arguments and by giving priority to the last introduced argument.
- We define the notion of appeal of an argument in terms of the goals which the revision triggered by the argument allows to satisfy by means of a plan.

It would be interesting to investigate how the work by Hunter [7] relates with conventional wisdom and our definition of appeal. Note that appeal must not be confused with wishful thinking: the receiver does not prefer a state of the world which makes its goals true, but one which gives him more opportunities to act to achieve its goals. The rationality of this kind of reasoning is discussed, e.g., by [1].

In this paper we do not study the formal properties of argumentative belief revision, and we do not relate it to the AGM postulates. However, from the paper it already appears that postulates like success are not meaningful in this framework. Moreover, we do not study how the different types of argumentation frameworks impact on belief revision [9].

References

1. G. Boella, C. da Costa Pereira, A. Tettamanzi, G. Pigozzi, and L. van del Torre. What you should believe: Obligations and beliefs. In *Proceedings of KI07-Workshop on Dynamics of Knowledge and Belief*, 2007.
2. G. Boella, C. D. C. Pereira, G. Pigozzi, A. Tettamanzi, and L. van der Torre. Choosing your beliefs. In G. Boella, L. W. N. van der Torre, and H. Verhagen, editors, *Normative Multi-agent Systems*, volume 07122 of *Dagstuhl Seminar Proceedings*. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.
3. C. da Costa Pereira and A. Tettamanzi. Towards a framework for goal revision. In *BNAIC 2006*, pages 99–106. University of Namur, 2006.
4. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*-person games. *Artificial Intelligence*, 77(2):321–358, 1995.
5. J. K. Galbraith. *The Affluent Society*. Houghton Mifflin, Boston, 1958.
6. S. O. Hansson. Logic of belief revision. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2006.
7. A. Hunter. Making argumentation more believable. In *AAAI 2004*, pages 269–274, 2004.
8. F. Paglieri and C. Castelfranchi. Revising beliefs through arguments: Bridging the gap between argumentation and belief revision in MAS. In I. Rahwan, P. Moraitis, and C. Reed, editors, *ArgMAS 2004*, volume 3366 of *Lecture Notes in Computer Science*, pages 78–94. Springer, 2005.
9. N. D. Rotstein, A. J. Garcia, and G. R. Simari. From desires to intentions through dialectical analysis. In *AAMAS '07*, pages 1–3, New York, NY, USA, 2007. ACM.