The Knowledge Engineering Review, Vol. 35, e36, 1 of 27.
doi:10.1017/S0269888920000405

Measuring the strength of threats, rewards, and appeals in persuasive negotiation dialogues

MARIELA MORVELI-ESPINOZA[1], JUAN CARLOS NIEVES[2], and CESAR AUGUSTO TACLA[1]

[1]*Program in Electrical and Computer Engineering (CPGEI), Federal University of Technology, Paraná (UTFPR), Curitiba, Brazil*
e-mails: morveli.espinoza@gmail.com; tacla@utfpr.edu.br
[2]*Department of Computing Science of Umeå University, Umeå, Sweden*
e-mail: jcnieves@cs.umu.se

Abstract

The aim of this article is to propose a model for the measurement of the strength of rhetorical arguments (i.e., threats, rewards, and appeals), which are used in persuasive negotiation dialogues when a proponent agent tries to convince his opponent to accept a proposal. Related articles propose a calculation based on the components of the rhetorical arguments, that is, the importance of the goal of the opponent and the certainty level of the beliefs that make up the argument. Our proposed model is based on the pre-conditions of credibility and preferability stated by Guerini and Castelfranchi. Thus, we suggest the use of two new criteria for the strength calculation: the credibility of the proponent and the status of the goal of the opponent in the goal processing cycle. We use three scenarios in order to illustrate our proposal. Besides, the model is empirically evaluated and the results demonstrate that the proposed model is more efficient than previous works of the state of the art in terms of numbers of negotiation cycles, number of exchanged arguments, and number of reached agreements.

1 Introduction

Negotiation is a key form of interaction, among agents, that is used for resolving conflicts and reaching agreements. Argumentation has been used in some works of negotiation because it allows an agent to exchange additional information, which can be used for justifying his proposals (e.g., Sierra *et al*. 1998; Amgoud *et al*. 2000; Rahwan *et al*. 2003; Dimopoulos & Moraitis 2011. Arguments used in negotiation dialogues are generally explanatory ones and allow agents to argue about their beliefs or other mental attitudes during the negotiation process (Rahwan *et al*. 2003). Nevertheless, there are other types of arguments that may act as persuasive elements. These ones are called rhetorical arguments[1] and are the following:

- *Threats*, which carry out sanctions when the opponent does not accept the proposal sent by the proponent.
- *Rewards*, which are used when the proponent wants to entice an opponent to do a certain action by offering to do another action as a reward or by offering something that the opponent needs.

† This is an extended version of the article accepted to be published in the Proceedings of the 17th European Conference on Multi-Agent Systems (Morveli-Espinoza *et al*. 2020).

[1] When an agent uses rhetorical arguments to back their proposals, the negotiation is called persuasive negotiation (Ramchurn *et al*. 2003).

- *Appeals*, which try to persuade the opponent by offering a reward; however, this recompense is not a consequence of an action of the proponent. If the proponent does not have a recompense to offer, he can appeal to one goal of the opponent that does not need the proponent's intervention. Appeals can be seen as self-rewards (Amgoud & Prade, 2004).

Let us consider a scenario of a Consumer Complaint Website whose goal is to try to resolve a conflict between consumers and companies. In this scenario, a software agent (denoted by CONSUMER) complains about a service on behalf of a human user and another software agent who acts on behalf of a company (denoted by COMPANY) offers possible solutions. In the following example, the user of CONSUMER missed an international flight due to a schedule change and he wants the airline company to reimburse him the total price of the ticket; however, the airline company only wants to refund the 20% of the total price of the ticket. At this point, CONSUMER tries to force COMPANY to accept his proposal and decides to send a threat. The following are three threats that the CONSUMER can generate:

- th_1 : *You should refund the total price of the ticket, otherwise I will never buy a ticket in your company anymore, so you will not reach your financial goals.*
- th_2 : *You should refund the total price of the ticket, otherwise I will destroy your reputation in social networks, so you will not gain the award to the Best Airline Frequent Flier Loyalty Program (BAFFLP).*
- th_3 : *You should refund the total price of the ticket, otherwise I will take legal actions against your company.*

The question is: which of these threats (arguments) will CONSUMER choose to try to persuade COMPANY to accept his proposal? According to Guerini and Castelfranchi (2006), a rhetorical argument has to meet some pre-conditions in order for the proponent to reach a negotiation favorable to him; therefore, the chosen argument has to be in the set of arguments that meet such pre-conditions. However, before the proponent decides what argument to send, he needs to have a way of differentiating the arguments of that set. A way of doing it is by measuring their strengths (Ramchurn *et al*. 2003). Thus, the research question of this article is: *What criteria should an agent take into account in order to measure the strength of a rhetorical argument and how should this measurement be done?*

There are few researches about the measurement of the strength of logic-based rhetorical arguments. Amgoud and Prade (2004, 2005a, 2006) take into account two criteria, namely (i) the importance of the opponent's goal and (ii) the certainty level of the beliefs that make up the argument. However, there exist situations in which other criteria are needed in order to perform a more exact measurement of the arguments strength. To make this discussion more concrete, consider the following situations:

- CONSUMER knows that 'reaching the financial goals' (denoted by go_1) and 'gaining the award to the BAFFLP' (denoted by go_2) are two goals of COMPANY—the opponent—that have the same importance. If CONSUMER only considers the importance of the opponent's goal to calculate the strength of the threats built with these goals, he cannot decide which threat to send because they have the same strength. Thus, there exists the need of another criterion—related to the COMPANY's goals—that helps CONSUMER to break the tie. In order to achieve a goal, it has to pass by some states. For instance, assume that go_1 has already been achieved; hence, threatening this goal would not be useful for CONSUMER. On the other hand, COMPANY has not yet achieved go_2; hence, attacking it can make COMPANY lose the award; and consequently, he will not achieve go_2.
- CONSUMER has already used rhetorical arguments with other companies before and rarely he has fulfilled what has been agreed, and agent COMPANY knows about it. In this case, the strength of a rhetorical argument sent by CONSUMER is also influenced by his credibility.

In the first case, notice that besides importance, there is another criterion to evaluate the worth of an opponent's goal, because it does not matter how important a goal is if it is far from being achieved or if it is already achieved. In the second case, the credibility of the proponent should also be considered,

since even when an opponent's goal is very important and/or achievable, a low level of credibility could impact on the strength value of an argument. Thus, the new suggested criteria for the measurement of the strength of rhetorical arguments are the proponent's credibility and the status of the opponent's goal.

To determine the possible statuses of a goal, we adopted the Belief-based Goal Processing (BBGP) model (Castelfranchi & Paglieri 2007). In this model, the processing of goals is divided in four stages: (i) activation, (ii) evaluation, (iii) deliberation, and (iv) checking; and the statuses a goal can adopt are: (i) active (=desire), (ii) pursuable, (iii) chosen (= future-directed intention), and (iv) executive (= present-directed intention). The status of a goal changes when it passes from one stage to the next. Thus, when it passes the activation stage it becomes active, when it passes the evaluation stage it becomes pursuable, and so on. A goal is closer to be achieved when it is closer of passing the last stage. Besides, we consider the cancelled status. A goal can be cancelled in every stage and the agent ceases to pursue it.

This work is an extended version of Morveli-Espinoza *et al*. (2020), being the main differences:

- Guerini and Castelfranchi (2006) claim that two pre-conditions should be fulfilled in order to determine convincing arguments. In Morveli-Espinoza *et al*. (2020), we assume that an agent does not differentiate between convincing and non-convincing arguments. Nevertheless, this is important because depending on the type of scenario the agent is negotiating, he will be able to determine or not convincing arguments. Thus, in this article we distinguish between fully and partially informed scenarios, which influences on the set of arguments the agent will use during the negotiation.
- In Morveli-Espinoza *et al*. (2020), we only propose one way for calculating the strength. In this work, we propose one way more, considering the data the agent has about his opponent.
- Two additional experiments were executed in order to evaluate the performance of the model in fully and partially informed scenarios and for comparing both ways of calculating the strength.
- Finally, the proposal was applied to three different scenarios.

This article is structured as follows: Section 2 presents the knowledge representation and the architecture of a BBGP-based agent. Section 3 is devoted to the logical definition of rhetorical arguments. Section 4 presents the strength calculation model. It includes the analysis of the criteria that will be considered and the dynamics of the model. Section 5 focuses on the application of the model to fully and partially informed scenarios. Section 6 presents the empirical evaluation of the proposed model. In Section 7, we discuss the related work. Finally, Section 8 summarizes this article and outlines future work.

2 Knowledge representation and negotiating agents

We use rule-based systems to represent the mental states of the agent. Thus, let $\mathcal{L}$ be a set of finite literals (literals are atoms or negation of atoms, the negation of an atom A is denoted by $\neg A$) $l, l_1, ..., l_n$ in first-order logical language and $\mathcal{C}$ is a set of finite constant symbols. Facts are elements of $\mathcal{L}$ and rules are of the form $r = l_1, ..., l_n \rightarrow l$. $\texttt{HEAD}(r) = l$ denotes the head of a rule and $\texttt{BODY}(r) = \{l_1, ..., l_n\}$ denotes the body of the rule. We assume that the body of every rule is finite and not empty. We now define a theory as a triple $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ where $\mathcal{F} \subseteq \mathcal{L}$ is a set of facts, $\mathcal{S}$ is a set of strict rules[2], and $\mathcal{D}$ is a set of defeasible rules. As a consequence operator, we use derivation schemas. The following definition was extracted from Amgoud & Besnard (2013).

DEFINITION 1 (Derivation schema). *Let $\mathcal{T} = \langle \mathcal{F}, \mathcal{S}, \mathcal{D} \rangle$ be a theory and $l \in \mathcal{L}$. A derivation schema for l from $\mathcal{T}$ is a finite sequence $T = \{(l_1, r_1), ..., (l_n, r_n)\}$ such that:*

- $l_n = l$, *for* $i = 1...n$,
  - $l_i \in \mathcal{F}$ *and* $r_i = \emptyset$, *or*
  - $r_i \in \mathcal{S} \cup \mathcal{D}$ *and* $\texttt{HEAD}(r_i) = l_i$ *and* $\texttt{BODY}(r_i) \subseteq \{l_1, ..., l_{i-1}\}$
- $\texttt{SEQ}(T) = \{l_1, ..., l_n\}$

[2] Strict rules are rules in classical sense, that is, the conclusion follows every time the antecedents hold whereas defeasible rules can be defeated by contrary evidence (Lam & Governatori 2011).

- $\texttt{FACTS}(T) = \{l_i \mid i \in \{1, ..., n\}, r_i = \emptyset\}$
- $\texttt{STRICT}(T) = \{r_i \mid i \in \{1, ..., n\}, r_i \in \mathcal{S}\}$
- $\texttt{DEFE}(T) = \{r_i \mid i \in \{1, ..., n\}, r_i \in \mathcal{D}\}$
- $\texttt{CN}(\mathcal{T})$ *denotes the set of all literals that have a derivation schema from* $\mathcal{T}$*, that is, the consequences drawn from* $\mathcal{T}$.

T is called minimal when $\nexists \mathcal{T}' \subset (\texttt{FACTS}(T), \texttt{STRICT}(T), \texttt{DEFE}(T))$ *such that* $l \in \texttt{CN}(\mathcal{T}')$. *A set* $\mathcal{L}' \subseteq \mathcal{L}$ *is called consistent iff* $\nexists l, l' \in \mathcal{L}'$ *such that* $l = \neg l'$*; otherwise, it is inconsistent.*

We can define now the architecture of a negotiating BBGP-based agent. It is a tuple $\langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \mathcal{A}_{val}, \texttt{REP}\rangle$ such that:

- $\mathcal{T}$ is the theory of the agent;
- $\mathcal{G}$ is the finite set of goals of the agent, whose elements are literals of $\mathcal{L}$;
- $\mathcal{O}pp$ is the finite set of opponents of the agent, whose elements are constants of $\mathcal{C}$;
- $\mathcal{GO} = \mathcal{GO}_a \cup \mathcal{GO}_p \cup \mathcal{GO}_c \cup \mathcal{GO}_e \cup \mathcal{GO}_{canc}$ is the finite set of the opponent's goals such that $\mathcal{GO}_a$ is the set of the active opponent's goals, $\mathcal{GO}_p$ is the set of the pursuable ones, $\mathcal{GO}_c$ is the set of the chosen ones, $\mathcal{G}_e$ is the set of the executive ones, and $\mathcal{GO}_{canc}$ is the set of the cancelled ones. These sets are pairwise disjoint.
- $\mathcal{S}_{\mathcal{O}pp}$ is a finite set of tuples $(op, \texttt{THRES}, L_{\mathcal{GO}})$ such that $op \in \mathcal{O}pp$, $\texttt{THRES} \in [0, 1]$ is the value of the threshold of the opponent[3], and $L_{\mathcal{GO}} = TH_{\mathcal{GO}} \cup RW_{\mathcal{GO}} \cup AP_{\mathcal{GO}}$ is the set of goals of opponent op such that these goals can be threatenable ($go \in TH_{\mathcal{GO}}$), rewardable ($go \in RW_{\mathcal{GO}}$), or appealable ($go \in AP_{\mathcal{GO}}$). It holds that $\forall go \in L_{\mathcal{GO}}$, $go \in \mathcal{GO}$, this means that if a goal is in the goals' list of an opponent—$L_{\mathcal{GO}}$—it is also in the opponent's goal set $\mathcal{GO}$. It also holds that $TH_{\mathcal{GO}}$, $RW_{\mathcal{GO}}$, and $AP_{\mathcal{GO}}$ are pairwise disjoint. Finally, let $\texttt{TH\_GO}(op) = TH_{\mathcal{GO}}$, $\texttt{RW\_GO}(op) = RW_{\mathcal{GO}}$, and $\texttt{AP\_GO}(op) = AP_{\mathcal{GO}}$ be three functions that return the sets of threatenable, rewardable, and appealable goals of op, respectively.
- $\mathcal{S}_{\mathcal{GO}}$ is a set of pairs $(go, \texttt{IMP})$ such that $go \in \mathcal{GO}$ and $\texttt{IMP} \in [0, 1]$ represents the importance value of go.
- $\mathcal{A}$ is the base of the proponent's actions, whose elements are ground atoms.
- $\mathcal{AO}$ is the base of the opponent's actions, whose elements are ground atoms. The role of action in our calculation model will be further explained below.
- $\mathcal{A}_{val}$ is a set of pairs (ac, val) such that $ac \in \mathcal{A}$ or $ac \in \mathcal{AO}$ is an action and $val \in [0, 1]$ is a real number that represents the value of action ac, where zero means that ac is not valuable at all whereas one is the maximum value of an action. Let $\texttt{VALUE}(ac) = val$ be a function that returns the value of a given action ac.
- $\texttt{REP} \in [0, 1]$ is the reputation value of the proponent, which is visible for any other agent.

When a proponent agent employs a rhetorical argument to try to convince an opponent to do a certain action, it can be seen as a goal of him. $\mathcal{G}$ is a set and the idea is that it can be divided in two sub-sets. What about: For this reason, one can distinguish two kinds of goals in $\mathcal{G}$: (i) goals that the agent himself has to perform actions to achieve them, and (ii) goals that need the opponent involvement to be achieved, for example, the goal of agent CONSUMER is that agent COMPANY ewfunds the total price of the ticket. For this goal to be achieved, it is necessary that COMPANY executes the required action. This type of goal is called *outsourced*.

DEFINITION 2 (Outsourced goal). *An outsourced goal g is an expression of the form g(op, ac), such that,* $op \in \mathcal{O}pp$ *and* $ac \in \mathcal{AO}$ *represents an action that op has to perform. Let* $\texttt{OPP}(g) = op$ *and* $\texttt{ACT}(g) = ac$ *be the functions that return each component of the outsourced goal g.*

We assume that a negotiating agent has in advance the necessary information for generating rhetorical arguments and for calculating their strengths. This information is related to the opponent's goals, the

[3] The threshold is a value used in the strength calculation model. This is better explained in Section 4.

status of these goals, the opponent's actions, and the values of these actions. In order to obtain such information, agent can gather information about his opponent(s). This approach is known as opponent modeling[4].

Baarslag *et al*. (2016) present a survey about some techniques of opponent modeling that are based on learning. Such techniques include Bayesian learning, non-linear regression, kernel density estimation, and artificial neural networks. Other works about opponent modeling with focus on argumentation are Hadjinikolis *et al*. (2013), Rienstra *et al*. (2013), Hadjinikolis *et al*. (2015), and Hunter (2015).

3 Threats, rewards, and appeals

In this section, we present the logical definitions of the rhetorical arguments that are being studied in this article.

3.1 Threats

The use of threats is a well-known strategy in negotiation (e.g., Sycara 1990; Sierra *et al*. 1998; Ramchurn *et al*. 2003); however, unlike rewards and appeals, threats have a negative nature.

Based on the three threats given in the example of the Introduction, we can say that a threat is mainly made up of two goals:

- **An opponent's goal:** It is the goal of the opponent that is being threatened by the proponent. It is a goal that the opponent wants to achieve or maintain. For example, '*maintaining a good reputation*', '*gaining customers fidelity*', and '*avoiding legal problems*'.
- **An outsourced goal of the proponent:** It is the goal of the proponent that needs the opponent involvement to be achieved. For example, '*getting that* `COMPANY` *refunds my ticket's money*'.

Following, we present the formal definition of a threat. This is based on the definition given in Amgoud and Prade (2004), with some modifications that consider the mental states of the negotiating BBGP-based agent and the rule-based approach.

DEFINITION 3 (Threat). *Let $\mathcal{T}$ be the theory of a negotiating BBGP-based agent, $\mathcal{G}$ be his goals base, and $\mathcal{GO}$ be his opponent's goals base. A threat constructed from $\mathcal{T}$, $\mathcal{G}$, and $\mathcal{GO}$ is a triple $th = \langle T, g, go \rangle$, where:*

- $go \in \mathcal{GO}$ *and* $go \in \texttt{TH\_GO}(\texttt{OPP}(g))$,
- $g \in \mathcal{G}$,
- $T \cup \neg\texttt{ACT}(g)$ *is a derivation schema for* $\neg go$ *from* $\mathcal{T}$,
- $\texttt{SEQ}(T)$ *is consistent,*
- T *is minimal.*

Let us call T the support of the threat, g its conclusion, and go is the threatened goal.

According to this definition, a threat is constructed under the hypothesis that the proponent's goal will not be achieved, which has a negative effect not only on the proponent but also on the opponent. The negative effect on the proponent is obviously that he does not achieve a goal and the negative effect on

[4] The opponent modeling problem is a complex process in any strategic interaction between intelligent (human/software) agents. By making use of opponent modeling, it is possible to represent necessary information about the opponent, which may be used during the negotiation encounter. Opponent modeling can be performed either online or offline; it depends on the availability of past data. Regarding offline models, these are created before the negotiation starts by using previously obtained data from earlier negotiations. Whereas online models are constructed from knowledge that is gather during a single negotiation encounter. The existence or not of previous data about opponents changes the maintenance of opponent modeling profiles. In this sense, we believe we can use user's profiles (like in recommending systems) and goal recognition techniques for improving the performance of our proposal.

the opponent is that he will not achieve one of his goals either. Thus, both agents need each other to achieve their goals.

3.2 Rewards and appeals

Rewards and appeals are also used during a negotiation dialogue as positive persuasive elements (e.g., Sierra *et al*. 1997; Shi *et al*. 2006; Florea & Kalisz 2007; Ramchurn *et al*. 2007). Both rewards and appeals result in a clear benefit for the opponent agent.

Example 1. Let us recall the Consumer Complaint Website scenario. However, now suppose that COMPANY is trying to offer a reward to CONSUMER:

- rw_1 : *If you agree with the 20% of refund, we will give you 10 000 miles.*
- rw_2 : *If you agree with the 20% of refund, we will sell you an executive ticket for the price of an economic one for any national destination.*
- rw_3 : *If you agree with the 20% of refund, we will give you our service of assistance for elderly for free for any destination.*

We can construct rewards and appeals in the same way as we construct threats. This means that rewards and appeals are also based on an opponent's goal and on an outsourced goal of the proponent.

Below, we present the formal definition of rewards and appeals. This is also based on the definition given in Amgoud and Prade (2004), with the necessary modifications that consider the mental states of the negotiating BBGP-based agent.

DEFINITION 4 (Reward/Appeal). *Let $\mathcal{T}$ be the theory of a negotiating BBGP-based agent, $\mathcal{G}$ be his goals base, and $\mathcal{GO}$ be his opponent's goals base. A reward/appeal constructed from $\mathcal{T}$, $\mathcal{G}$, and $\mathcal{GO}$ is a triple $\langle T, g, go\rangle$, where:*

- $go \in \mathcal{GO}$,
- *For rewards:* $go \in \texttt{RW\_GO}(\texttt{OPP}(g))$ *and for appeals:* $go \in \texttt{AP\_GO}(\texttt{OPP}(g))$,
- $g \in \mathcal{G}$,
- $T \cup \texttt{ACT}(g)$ *is a derivation schema for go from* $\mathcal{T}$,
- $\texttt{SEQ}(T)$ *is consistent,*
- *T is minimal.*

Let us call T the support of the reward/appeal, g its conclusion, and go is the rewardable/appealable goal. Furthermore, let RHETARG *denote the set of threats, rewards, and appeals that an agent can construct from his theory* $\mathcal{T}$.

4 Strength calculation model

In this section, we start by analyzing the necessary criteria for evaluating the strength of threats, rewards, and appeals. Next we detail the steps the proponent agent follows in order to obtain the strength values of the arguments he generates.

4.1 Pre-conditions of credibility and preferability

According to Guerini and Castelfranchi (2006), a rhetorical argument has to meet some pre-conditions in order the proponent can reach a negotiation favorable to him. Consequently, the chosen rhetorical argument has to belong to the set of rhetorical arguments that meet such pre-conditions. These pre-conditions are related to the credibility of the proponent agent and to the preferability of the opponent's goal regarding the requested action.

4.1.1 Credibility

According to Guerini and Castelfranchi (2006), Castelfranchi and Guerini (2007), there exists a goal cognitive structure when a proponent utters an influencing sentence to an opponent such that the first goal of the proponent agent is related to his credibility. In other words, when a proponent agent wants to persuade an opponent agent, the opponent has to believe that he (the proponent) is credible.

In this work, in order to evaluate the credibility of the proponent, we take into account the following concepts:

1. **The proponent's reputation:** Reputation can be defined as a social notion associated with how trustworthy an individual is within a society. The estimate value of reputation is formed and updated over time with the help of different sources of information. Several computational models of reputation consider that reputation can be estimated based on two different sources: (i) the direct interactions and (ii) the information provided by other members of the society about experiences they had in the past (e.g., Yu & Singh 2000; Sabater & Sierra 2001; Pinyol & Sabater-Mir 2013). Another works about trust and reputation are Falcone and Castelfranchi (2001, 2004).

   In this work, reputation can be seen as the 'social' notion—within an agents society—about how trustworthy the proponent is with respect to fulfil his threats, rewards, and appeals. In other words, it is an evidence of the proponent's past behavior with respect to his opponents. We assume that this value is already estimated and it is not private information. Thus, reputation value of the proponent is known by any other agent. It means that when the proponent begins a negotiation with other agent (his opponent), this one is conscious of the reputation of the proponent. We also assume that the proponent has only one reputation value for the three kinds of rhetorical arguments.

   The reputation value of a proponent agent P is represented by a real number: $\texttt{REP}(P) \in [0, 1]$ where zero represents the minimum reputation value and one the maximum reputation value.
2. **The opponent's credibility threshold:** It is used to indicate the lowest value of the proponent's reputation so that the opponent considers a rhetorical argument credible. Thus, the credibility threshold of an opponent agent O is represented by a real number: $\texttt{THRES}(O) \in [0, 1]$ where zero represents the minimum threshold value and one the maximum threshold value.

   A low threshold denotes a trusting (or easier to be persuaded) opponent and a high threshold denotes a mistrustful opponent, that is, more difficult to be persuaded. We assume that the proponent knows the values of the thresholds of his possible opponents and stores these values.

The proponent evaluates his own credibility—in the eyes of his opponent—by comparing both values: the proponent's reputation and the opponent's threshold. When the reputation value is greater than or equal to the opponent's threshold, it means that the proponent believes that the opponent considers him (the proponent) credible; otherwise, the opponent believes that the proponent is not credible.

DEFINITION 5 (Proponent's credibility). *Let P be a proponent agent, $\texttt{REP}(P)$ be his reputation, and $\texttt{THRES}(O)$ be the threshold of his opponent O. P is credible if $\texttt{REP}(P) \geqslant \texttt{THRES}(O)$; otherwise, O does not believe that P is credible.*

4.1.2 Preferability

The second pre-condition a proponent agent has to meet in order to attain a favorable negotiation is the *preferability* (Guerini & Castelfranchi 2006). This pre-condition is based on the relation between the opponent's goal and the action he is required to perform. According to Guerini and Castelfranchi (2006), the opponent's goal must be more valuable for him (the opponent) than performing the required action.

We first present the criteria that will be evaluated in order to estimate how valuable a goal is for the opponent. Below, we analyze each criteria and indicate how the value of the opponent's goal will be estimated.

1. **Importance of the opponent's goal:** It is related to how meaningful the goal is for the opponent. The value of the importance of a given goal go is a real number represented by: $\texttt{IMP}(go) \in [0, 1]$ where

zero means that the goal is not important at all, and one is the maximum importance of the goal. The more important a goal is for the opponent, the more threatenable, rewardable, or appealable this goal is.

2. **Effectiveness of the opponent's goal:** It is related to the degree to which an opponent's goal is successful for persuasion and it is based on the status of the goal in the intention formation process. Let us recall that we are working with BBGP-based agents; therefore, the goals base of the opponent is divided into five sub-sets: active goals, pursuable goals, chosen goals, executive goals, and cancelled goals. A goal is close of being achieved when its status is chosen or executive and it is far of being achieved when its status is active or pursuable. Thus, depending on its status, a goal can be considered more or less threatenable, rewardable, or appealable. Let us analyze each case:
   - **Threatenable goal:** Recall that threats have a negative nature. In terms of the status of a goal, it means that a threat may make a goal go back to a previous status. In this work, we assume that every threatened goal will become cancelled. Therefore, a goal is considered more threatenable when its status is executive and less threatenable when its status is active. This is because an agent has more to lose when an executive goal is threaten than when an active goal is threaten. Regarding a cancelled goal, it is not threatenable at all.
   - **Rewardable and appealable goal:** In this case, both rewards and appeals have a positive nature. In terms of the status of a goal, it means that a reward/appeal may make a goal go forward to an advanced status. In this work, we assume that every rewarded/appealed goal will become executive. Therefore, a goal is considered more rewardable/appealable when its status is cancelled and less rewardable/appealable when its status is chosen. This is because an agent has more to win when a cancelled goal is rewarded/appealed than when a chosen goal is rewarded/appealed. Executive goals cannot be rewarded/appealed because the proponent has nothing to offer that makes them go forward. Therefore, executive goals are not rewardable/appealable at all.

The value of the effectiveness of a goal go depends on the argument that is built from it. We denote by $\texttt{arg}(go) \in \{th, rw, ap\}$ the type of argument that can be built where th means that the type of argument is a threat, rw means that the type of argument is a reward, and ap means that the type of argument is an appeal. The effectiveness of an opponent's goal go is represented by $\texttt{eff}(go) \in \{0, 0.25, 0.5, 0.75, 1\}$ such that zero means that go is not effective at all and one means that go is completely effective. The effectiveness of an opponent's goal is evaluated as follows:

$$\texttt{EFF}(go) = \begin{cases} 0 & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{canc}, \text{ or} \\ & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{e}, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{e} \\ \hline 0.25 & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{a}, \text{ or} \\ & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{c}, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{c} \\ \hline 0.5 & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{p}, \text{ or} \\ & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{p}, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{p} \\ \hline 0.75 & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{c}, \text{ or} \\ & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{a}, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{a} \\ \hline 1 & \text{if } \texttt{arg}(go) = th \text{ and } go \in \mathcal{GO}_{e}, \text{ or} \\ & \text{if } \texttt{arg}(go) = rw \text{ and } go \in \mathcal{GO}_{canc}, \text{ or} \\ & \text{if } \texttt{arg}(go) = ap \text{ and } go \in \mathcal{GO}_{canc} \end{cases}$$

Based on the importance and the effectiveness of a opponent's goal, it can be estimated how valuable this goal is. Thus, the worth of an opponent's goal is represented by $\texttt{WORTH}: \mathcal{GO} \rightarrow [0, 1]$ and it is estimated as follows.

DEFINITION 6 (Worth of the opponent's goal). *Let go be an opponent's goal,* $\texttt{IMP}(go)$ *its importance, and* $\texttt{EFF}(go)$ *its effectiveness. The equation for calculating the worth of go is*

$$\texttt{WORTH}(go) = \frac{\texttt{IMP}(go) + \texttt{EFF}(go)}{2} \tag{1}$$

We use the average value in order to obtain the final value of the worth of an opponent's goal because we consider that both criteria are equally significant to make the calculation and they do not overlap each other, since each of them characterizes a different aspect of the goal. We also want to keep the values of the worth of the goal in the same interval than the values of the both criteria, namely importance and effectiveness.

So far, we have analyzed the criteria to estimate how valuable an opponent's goal is. In order to evaluate the pre-condition preferability, the proponent should also know the value the opponent gives to the required action in order to compare both values. If the value of the opponent's goal is greater than the value of the required action then, the argument that uses that goal is considered preferable. Let us explain it with human examples. During an assault, the thief threatens the victim with the following sentence: '*If do not give me your bag, I hurt you*'. In this situation, it is rational to think that the physical well-being is above all. Hence, the value of the goal of the victim (the opponent) is greater than the value of the required action. In another scenario, we have a boss who is trying to convince one of his employees to work on Saturdays with the following reward: '*If you work every Saturday, then I give you a chocolate*'. In this situation, it is reasonable to believe that the value of the opponent's goal is not greater than the value of the required action.

Guerini and Castelfranchi (2006) claim that a threat, reward, or appeal can be considered convincing when it is credible and preferable. However, depending on the scenario, the proponent agent may or not have information about the real value of an action for his opponent. Thus, we can divide the scenarios into: (i) *fully informed scenarios*, in which the proponent knows both the value of the actions for his opponent and the value of the opponent's goals; and (ii) *partially informed scenarios*, in which the proponent only know the value of the opponent's goals. Therefore, the preferability of a given goal can only be evaluated in fully informed scenarios. An example of this kind of scenario may be a robots scenario, where the agents share the same actions and values for that actions. In partially informed scenarios, the preferability cannot be evaluated; however, the proponent agent has the value of the opponent's goal. While it is true that the proponent will not know if an argument is convincing, he can base on the value of its strength to decide which argument to choose and send to his opponent.

For fully informed scenarios, the preferability of a given opponent's goal is determined in the following definition.

DEFINITION 7 (Preferability of an opponent's goal). *Let* $go \in \mathcal{GO}$ *be an opponent's goal and* $ac \in \mathcal{AO}$ *an opponent's action. Goal go is preferable if* $\texttt{WORTH}(go) > \texttt{VALUE}(ac)$*; otherwise, it is not preferable.*

Notice that in the rescue robots scenario, the base of actions may be the same for all of them. Therefore, we may have that $\mathcal{A} = \mathcal{AO}$.

4.2 Steps of the model

In the previous sub-section, we have studied the pre-conditions for a rhetorical argument to be considered convincing. In other words, if a rhetorical argument meets the pre-conditions previously presented, then the proponent agent believes that he is able to convince his opponent to perform the requested action. Assuming that the agent has more than one convincing rhetorical argument, he still needs a way to compare such arguments. Therefore, a strength value for each argument is still necessary. Thus, in this sub-section we will study how these pre-conditions can be combined to obtain the strength of each argument. We propose a strength calculation model based on the previously studied pre-conditions and that

can be applied when the proponent has only one possible opponent or when the proponent can choose an opponent to send an argument. The output of this proposal is a set of rhetorical arguments with their respective strength values.

The first step of the calculation model is related to the proponent's credibility. Let us recall that an opponent agent has a credibility threshold that indicates the lowest value of the reputation of the proponent agent so the arguments of him may be considered credible. Let us suppose that a proponent agent P—with reputation value $\texttt{REP}(P) = 0.6$—has two opponents O_1 and O_2 and let $\texttt{THRES}(O_1) = 0.5$, $\texttt{THRES}(O_2) = 0.8$ be the thresholds of agents O_1 and O_2, respectively. Since $\texttt{REP}(P) \geqslant \texttt{THRES}(O_1)$, the proponent P can continue in the process of evaluation of the strength of the arguments addressed to O_1, but it does not occur for agent O_2 because $\texttt{REP}(P) \ngeq \texttt{THRES}(O_2)$.

When the proponent is considered credible by his opponent(s), the next step of the model is related to the preferability notion. It is important to highlight that only when the proponent believes that he is credible he can continue to the next step. Thus, regarding preferability two possibilities can be distinguished:

1. *Fully informed scenarios*: We can differentiate two sets of arguments. One set includes the arguments that are constructed using a preferable opponent's goal and the other set includes the arguments that are constructed using non-preferable goals. All the arguments of the first set are considered convincing. This means that any of these arguments can make the opponent performs the required action. The strength value of these arguments is considered an *absolute value*. Thus, in this kind of scenarios the agent is sure that will convince his opponent if the first set has at least one argument.
2. *Partially informed scenarios*: In this kind of scenario, we only have one set of arguments whose strength values are considered *relative values* because the agent is not sure about the convincing power of his arguments.

We can notice that the preferability concept has a direct impact on the assurance of the agent that his arguments are convincing or not. However, the preferability itself is not a value that represents the strength of an argument. Since the preferability evaluation is based on the worth of the opponent's goal, we will use this value in order to rank the arguments addressed to a given opponent. Thus, we will call this value the *basic strength* of an argument.

DEFINITION 8 (Basic strength). *Let $A = \langle T, g, go\rangle$ be a rhetorical argument; the basic strength of A is obtained by applying the same formula used for calculating the worth of go.*

$$\texttt{ST\_BASIC}(A) = \texttt{WORTH}(go) = \frac{\texttt{IMP}(go) + \texttt{EFF}(go)}{2} \tag{2}$$

A direct consequence of the above definition is that the value of the basic strength of a rhetorical argument is a real number between 0 and 1. Formally:

Property 1. Let $A = \langle T, g, go\rangle$ be a rhetorical argument. Since the value of the importance of $go \in [0, 1]$ and the effectiveness value of go are also between 0 and 1, then $\texttt{ST\_BASIC}(A) \in [0, 1]$, where 0 represents the minimum value and 1 represents the maximum value of the basic strength.

The basic strength is useful for ranking the arguments; however, for a more accurate value of the strength of the arguments, we can also take into account the credibility value. We will call this value the *combined strength* of an argument.

Before presenting the formula for calculating the combined strength of an argument, let us analyze the following situation. P is a proponent agent and O his opponent, let $\texttt{REP}(P) = 0.6$ be the reputation of agent P, and $\texttt{THRES}_1(O) = 0.5$ and $\texttt{THRES}_2(O) = 0.2$ be two possible thresholds of O. We can notice that $\texttt{THRES}_1$ reflects a less credulous attitude than $\texttt{THRES}_2$; thus, although P is credible in both cases, the 'accurate' value of P's credibility is different for each case since the difference between $\texttt{REP}(P)$ and $\texttt{THRES}_1$ is less than the difference between $\texttt{REP}(P)$ and $\texttt{THRES}_2$. Therefore, the credibility value of P should have an impact on the calculation of the strength of the arguments because the higher

the difference between the threshold value and the reputation value is, the higher the credibility of the proponent is.

We use next equation to calculate the 'accurate' value of the credibility of P with respect to an opponent O, whose threshold is $\texttt{THRES}(O)$.

$$\texttt{ACCUR\_CRED}(P, O) = \texttt{REP}(P) - \texttt{THRES}(O) \tag{3}$$

This value is used to obtain the combined strength of the arguments. Thus, the combined strength of an argument depends on the basic strength of the argument and the 'accurate' value of the proponent's credibility.

DEFINITION 9 (Combined strength). *Let $A = \langle T, g, go\rangle$ be a rhetorical argument and $O \in \mathcal{O}pp$ be an opponent whose threatened/rewarded/appealed goal is go. The combined strength of A is obtained by applying:*

$$\texttt{ST\_COMB}(A) = \texttt{ST\_BASIC}(A) \times \texttt{ACCUR\_CRED}(P, O) \tag{4}$$

We can say that the value of the combined strength of an argument is a real number that is between zero and the product of the basic strength times the proponent reputation value. Thus, the combined strength has its maximum value when the opponent's threshold is zero and the basic strength of the argument is maximal. Formally:

Property 2. Let $A = \langle T, g, go\rangle$ be a rhetorical argument—whose basic strength is $\texttt{ST\_BASIC}(A)$—and $\texttt{REP}(P)$ be the value of the proponent's reputation. The combined strength of A is a real number that is in the following interval $\texttt{ST\_COMB}(A) \in [0, \texttt{ST\_BASIC}(A) \times \texttt{REP}(P)]$.

Figure 1 depicts a workflow of our approach for the strength evaluation. In summary, first of all, the proponent has to evaluate his credibility with respect to his opponent, if he is not credible enough he stops the process; otherwise, he continues. Depending on the type of scenario, the preferability is evaluated or not. Besides, the agent may choose to take into account the accuracy, in such case the combined strength is calculated; otherwise, he only calculates the basic strength.

5 Applying the calculation model to fully and partially informed scenarios

In this section, we present the application of the proposed model to three scenarios, the Consumer Complaint Website scenario, a rescue robots scenario, and the patients medication scenario, which will be described below. The first and the last scenarios are partially informed scenarios, whereas the second one is a fully informed scenario. We start presenting the logical formalization and then we make the strength calculations for each scenario.

5.1 Consumer Complaint Website scenario

In this scenario, we work with threats and rewards. Recall that this scenario is an example of partially informed scenario; therefore, the value of the arguments strength is considered relative.

THREATS

Next, we present the mental state of agent CONSUMER and the logical formalization of each threat. Hereafter, we omit some elements from the mental state because they are not necessary.

$\texttt{CONSUMER} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP}\rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$$\mathcal{S} = \{\neg\, refund(money) \rightarrow \neg buy\_again(ticket),$$
$$\neg\, refund(money) \rightarrow destroy(rep\_social\_net),$$

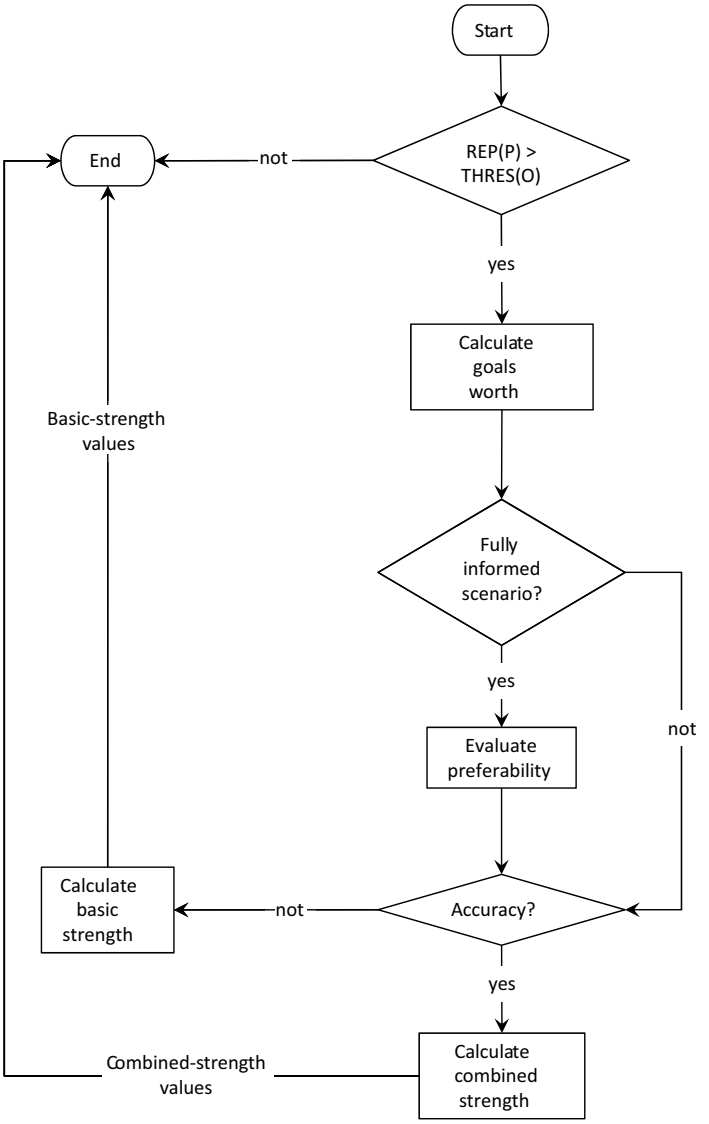

Figure 1 Workflow of the proposed strength calculation model

$$\neg\, refund(money) \rightarrow take(legal\_actions),$$
$$\neg buy\_again(ticket) \rightarrow \neg gain(costu\_fidel),$$
$$destroy(rep\_social\_net) \rightarrow \neg have(good\_rep),$$
$$take(legal\_actions) \rightarrow \neg avoid(legal\_probs)\}$$

$\mathcal{G} = \{g\}$ such that $g = get(\texttt{COMPANY}, \text{`}refund(money)\text{'})$

$\mathcal{O}pp = \{\texttt{COMPANY}\}$

$\mathcal{GO}_a = \{go_3\}$, $\mathcal{GO}_c = \{go_1\}$, $\mathcal{GO}_e = \{go_2\}$ such that $go_1 = gain(costu\_fidel)$, $go_2 = have(good\_rep)$, and $go_3 = avoid(legal\_probs)$

$\mathcal{S}_{\mathcal{O}pp} = \{(\texttt{COMPANY}, 0.75, \{go_1, go_2, go_3\})\}$ such that $\texttt{THRES}(\texttt{COMPANY}) = 0.75$, and $\{go_1, go_2, go_3\} \in TH_{\mathcal{GO}}$

$\mathcal{S}_{\mathcal{GO}} = \{(go_1, 0.85), (go_2, 0.85), (go_3, 0.7)\}$, and $\texttt{REP} = 0.8$.

Table 1. Strength values of the threats of agent CONSUMER in the software agents scenario

Goal	IMP(*go*)	Status	EFF(*go*)	ST_BASIC(*th*)	ST_COMB(*th*)	
go_1	0.85	Chosen	0.75	0.8	0.04	th_1
go_2	0.85	Executive	1	0.925	0.0463	th_2
go_3	0.7	Active	0.25	0.475	0.0238	th_3

From this mental state, the following threats can be generated:

$th_1 = \langle T_1, g, go_1 \rangle$ where:

$T_1 \cup \neg\text{SECOND}(g) = \{(\neg\ refund(money), \emptyset),$

$(\neg buy\_again(ticket), \neg\ refund(money) \rightarrow \neg buy\_again(ticket)),$

$(\neg gain(costu\_fidel), \neg buy\_again(ticket) \rightarrow \neg gain(costu\_fidel))\}$

$th_2 = \langle T_2, g, go_2 \rangle$ where:

$T_2 \cup \neg\text{SECOND}(g) = \{(\neg\ refund(money), \emptyset),$

$(destroy(rep\_social\_net), \neg\ refund(money) \rightarrow destroy(rep\_social\_net)),$

$(\neg have(good\_rep), destroy(rep\_social\_net) \rightarrow \neg have(good\_rep))\}$

$th_3 = \langle T_3, g, go_3 \rangle$ where:

$T_3 \cup \neg\text{SECOND}(g) = \{(\neg\ refund(money), \emptyset),$

$(take(legal\_actions), \neg\ refund(money) \rightarrow take(legal\_actions)),$

$(\neg avoid(legal\_probs), take(legal\_actions) \rightarrow \neg avoid(legal\_probs))\}$

According to the calculation model, firstly the credibility of CONSUMER has to be evaluated. Since REP(CONSUMER) > THRES(COMPANY) (i.e., $0.8 > 0.75$), we can proceed to calculate the basic strength values of the threats generated by CONSUMER. Since there is only one possible opponent, then it is not necessary to make the calculation of the combined strength values. Table 1 shows the basic and combined values of the strength of the threats generated by agent CONSUMER; the combined strength is calculated considering that ACCUR_CRED(CONSUMER, COMPANY) $= 0.8 - 0.75 = 0.05$.

Thus, we have that threat th_2—whose opponent's goal is *have*(*good_rep*)—is the strongest one and threat th_3—whose opponent's goal is *avoid*(*legal_probs*)—is the least strong threat.

REWARDS

Let us recall the rewards that agent COMPANY can construct to try to convince agent CONSUMER to accept only the 20% refund. If CONSUMER decides to send his strongest threat (i.e., threat th_2), COMPANY can construct a counter-threat to such threat. Thus, COMPANY would have both rewards and threats to support his position. In natural language, the rewards and threat that COMPANY can generate are

- rw_1 : *If you agree with the 20% of refund, we will give you 10 000 miles.*
- rw_2 : *If you agree with the 20% of refund, we will sell you an executive ticket for the price of an economic one for any national destination.*
- rw_3 : *If you agree with the 20% of refund, we will give you our service of assistance for elderly for free for any destination.*
- th_4 : *You should not destroy my reputation, otherwise I will denounce you for defamation and I will ask for a payment of civil reparations amounting to 1000 in favor of me.*

Next, we present the mental state of agent COMPANY and the logical formalization of the three rewards and the threat.

$\texttt{COMPANY} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$\mathcal{S} = \{accept(refund\_20) \rightarrow gain(miles),$
$accept(refund\_20) \rightarrow get\_discount(exec\_ticket),$
$accept(refund\_20) \rightarrow free(elderly\_assist),$
$\neg drop(destroy\_rep) \rightarrow make(denounce\_difam),$
$make(denounce\_difam) \rightarrow pay\_repar(1000),$
$pay\_repar(1000) \rightarrow \neg avoid(extra\_budget)\}$

$\mathcal{G} = \{g_1, g_2\}$ such that $g_1 = get(\texttt{CONSUMER}, \text{`}accept(refund\_20)\text{'})$ and

$g_2 = get(\texttt{CONSUMER}, \text{`}drop(destroy\_rep)\text{'})$

$\mathcal{O}pp = \{\texttt{CONSUMER}\}$

$\mathcal{GO}_a = \{go_6\}$, $\mathcal{GO}_c = \{go_4, go_7\}$, $\mathcal{GO}_{canc} = \{go_5\}$ such that $go_4 = gain(miles)$,

$go_5 = get\_discount(exec\_ticket)$, $go_6 = free(elderly\_assist)$, and $go_7 = avoid(extra\_budget)$

$\mathcal{S}_{\mathcal{O}pp} = \{(\texttt{CONSUMER}, 0.7, \{go_4, go_5, go_6, go_7\})\}$ such that $\texttt{THRES}(\texttt{CONSUMER}) = 0.7$,

$\{go_4, go_5, go_6\} \in RW_{\mathcal{GO}}$, and $\{go_7\} \in TH_{\mathcal{GO}}$

$\mathcal{S}_{\mathcal{GO}} = \{(go_4, 0.8), (go_5, 0.7), (go_6, 0.4), (go_7, 0.9)\}$, and $\texttt{REP} = 0.9$

From this mental state, the following rewards and threat can be generated:

$rw_1 = \langle T_1, g_1, go_4 \rangle$ where:

$T_1 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset), (gain(miles), accept(refund\_20) \rightarrow gain(miles))\}$

$rw_2 = \langle T_2, g_1, go_5 \rangle$ where:

$T_2 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset), (get\_discount(exec\_ticket), accept(refund\_20) \rightarrow get\_discount(exec\_ticket))\}$

$rw_3 = \langle T_3, g_1, go_6 \rangle$ where:

$T_3 \cup \texttt{SECOND}(g_1) = \{(accept(refund\_20), \emptyset), (free(elderly\_assist), accept(refund\_20) \rightarrow free(elderly\_assist))\}$

$th_4 = \langle T_4, g_2, go_7 \rangle$ where:

$T_4 \cup \neg\texttt{SECOND}(g_2) = \{(\neg drop(destroy\_rep), \emptyset),$
$(make(denounce\_difam), \neg drop(destroy\_rep) \rightarrow make(denounce\_difam))$
$(pay\_repar(1000), make(denounce\_difam) \rightarrow pay\_repar(1000))$
$(\neg avoid(extra\_budget), pay\_repar(1000) \rightarrow \neg avoid(extra\_budget))\}$

First of all, the credibility of COMPANY has to be evaluated. Since $\texttt{REP}(\texttt{COMPANY}) > \texttt{THRES}(\texttt{CONSUMER})$ (i.e., $0.9 > 0.7$), we can proceed to calculate the basic strength values of the arguments generated by COMPANY. Like in the previous case, since there is only one possible opponent, then it is not necessary to make the calculation of the combined strength values. Table 2 shows the basic and combined values of the strength of the rewards and the threat generated by agent COMPANY; the combined strength is calculated considering that $\texttt{ACCUR\_CRED}(\texttt{COMPANY}, \texttt{CONSUMER}) = 0.9 - 0.7 = 0.2$.

Thus, we have that reward rw_2—whose opponent's goal is $get\_discount(exec\_ticket)$—is the strongest rhetorical argument and reward rw_1—whose opponent's goal is $gain(miles)$—is the least strong argument. However, notice that the strength value of the unique threat is very close to the strength value of

Table 2. Strength values of the rewards and the threat of agent COMPANY in the software agents scenario

Goal	IMP(*go*)	Status	EFF(*go*)	ST_BASIC(*rw*)	ST_COMB(*rw*)	
go_4	0.8	Chosen	0.25	0.525	0.105	rw_1
go_5	0.7	Cancelled	1	0.85	0.17	rw_2
go_6	0.4	Active	0.75	0.575	0.115	rw_3
Goal	IMP(*go*)	Status	EFF(*go*)	ST_BASIC(*th*)	ST_COMB(*th*)	
go_7	0.9	Chosen	0.75	0.825	0.165	th_4

reward rw_2. This means that depending on the strategy of the agent, he can choose to send a threat or a reward.

5.2 Rescue robots scenario

This is a scenario of a natural disaster, where a set of robot agents have a set of tasks such as: (i) looking through rubble to find survivors, (ii) wandering the area in search of people needing help, (iii) helping disabled people do tasks, and (iv) bringing supplies for survivors. When they find a person who is seriously injured, the robots take him/her to the hospital, otherwise he/she is sent to a shelter. The robots can communicate with each other in order to ask for/send information or to ask for help.

The disaster area is divided into numbered zones, which are named by using ordered pairs. In the disaster area, there is also a robot workshop, where they can supply of power to keep working and be fixed, in case of a damage or failure.

Each agent is in charge of a certain zone and must achieve its own goals with respect to that zone, which are closely related to its tasks. However, robots can help each other in certain situations, for example, to remove heavy debris. It is under these conditions where a persuasive negotiation dialogue may arise, because robots have to decide whether to continue with their tasks and accomplish their own goals or stop to help another robot.

In this scenario, we work with appeals. Notice that this scenario is a sample of fully informed scenario; therefore, the concept of preferability is employed and the arguments strength is considered absolute.

Next, we present three appeals that a robot agent TOM can generate in order to try to convince another robot agent BOB to help him with a heavy debris.

- ap_1 : *If you help me, you can win utility points.*
- ap_2 : *If you help me, you can recharge your battery since the workshop is next to this zone.*
- ap_3 : *If you help me, you can fix your sensor since the workshop is next to this zone.*

Next, we present the mental state of agent TOM and the logical formalization of the three appeals. $\texttt{TOM} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$\mathcal{S} = \{help\_with(debris) \rightarrow gain(util\_points),$

$help\_with(debris) \rightarrow go(workshop),$

$go(workshop) \rightarrow recharge(battery),$

$go(workshop) \rightarrow fix(sensor),$

$\mathcal{G} = \{g_3\}$ such that $g_3 = get(\texttt{BOB}, \text{`}help\_with(debris)\text{'})$

$\mathcal{O}pp = \{\texttt{BOB}\}$

$\mathcal{GO}_a = \{go_9\}, \mathcal{GO}_p = \{go_8\}, \mathcal{GO}_c = \{go_{10}\}$ such that

$go_8 = gain(util\_points)$, $go_9 = recharge(battery)$, and $go_{10} = fix(sensor)$

Table 3. Strength values of the appeals of agent TOM in the rescue robots scenario

Goal	IMP(go)	Status	EFF(go)	ST_BASIC(ap)	ST_COMB(ap)	
go_8	0.7	Pursuable	0.5	0.6	0.06	ap_1
go_9	0.9	Active	0.75	0.825	0.083	ap_2
go_{10}	0.75	Chosen	0.25	0.5	0.05	ap_3

$\mathcal{S}_{\mathcal{O}pp} = \{(\text{BOB}, 0.7, \{go_8, go_9, go_{10}\})\}$ such that $\text{THRES}(\text{BOB}) = 0.7$, and $\{go_8, go_9, go_{10}\} \in AP_{\mathcal{GO}}$

$\mathcal{S}_{\mathcal{GO}} = \{(go_8, 0.7), (go_9, 0.9), (go_{10}, 0.75)\}$

$\mathcal{A} = \mathcal{AO} = \{help\_with(debris)\}$

$\mathcal{A}_{val} = \{(help\_with(debris), 0.55)\}$

$\text{REP} = 0.8$

From this mental state, the following appeals can be generated:

$ap_1 = \langle T_1, g_3, go_8 \rangle$ where:

$T_1 \cup \text{SECOND}(g_3) = \{(help\_with(debris), \emptyset), (gain(util\_points), help\_with(debris) \rightarrow gain(util\_points))\}$

$ap_2 = \langle T_2, g_3, go_9 \rangle$ where:

$T_2 \cup \text{SECOND}(g_3) = \{(help\_with(debris), \emptyset),$
$(go(workshop), help\_with(debris) \rightarrow go(workshop))$
$(recharge(battery), go(workshop) \rightarrow recharge(battery))\}$

$ap_3 = \langle T_3, g_3, go_{10} \rangle$ where:

$T_3 \cup \text{SECOND}(g_3) = \{(help\_with(debris), \emptyset),$
$(go(workshop), help\_with(debris) \rightarrow go(workshop))$
$(fix(sensor), go(workshop) \rightarrow fix(sensor))\}$

Like in previous scenarios, the credibility of TOM has to be evaluated. We have that TOM is considered credible by BOB based on $\text{REP}(\text{TOM}) > \text{THRES}(\text{BOB})$ (i.e., $0.8 > 0.7$). Thus, we can proceed to calculate the basic strength values of the appeals generated by TOM. Table 3 shows the basic and combined values of the strength of the appeals generated by agent TOM; the combined strength is calculated considering that $\text{ACCUR\_CRED}(\text{TOM}, \text{BOB}) = 0.8 - 0.7 = 0.1$.

Recall that $\text{ST\_BASIC}(ap) = \text{WORTH}(go)$ such that go is the opponent's goal that makes up the appeal ap. Thus, we can compare the values of the opponent's goals and the value of the required action. The result is the following:

$\text{WORTH}(go_8) > \text{VALUE}(help\_with(debris))$ $(0.6 > 0.55)$

$\text{WORTH}(go_9) > \text{VALUE}(help\_with(debris))$ $(0.825 > 0.55)$

$\text{WORTH}(go_{10}) \not> \text{VALUE}(help\_with(debris))$ $(0.5 < 0.55)$

Therefore, we have that appeals ap_1 and ap_2 are convincing ones, whereas appeal ap_3 is not convincing. This means that the set of arguments that agent TOM can use during the negotiation dialogue has been reduced to two.

5.3 Patients medication scenario

This is a scenario of a smart Medication Coach Intelligent Agent (MCIA) that supports patients in handling their medicine (Ingeson *et al.* 2018; Blusi & Nieves 2019). The MCIA manages different types

Figure 2 A MCIA used in a homelike environment. Extracted from Blusi and Nieves (2019)

of information such as the medication plan of the patients, medication restrictions, and the patient's preferences. Besides it perceives input data about the environment and the user activities through an AR[5] -headset (see Figure 2 for an illustration where a patient interacts with the MCIA through Microsoft HoloLens). The goal of the MCIA is to make sure that patients take their medicines at the times they are specified. With this aim, MCIA sends reminders to the patients; however, they can intentionally dismiss such reminders. It is in this point that rhetorical arguments can be used to try to convince patients to take their medicine. For example, assuming that a patient—let us call him John—needs to take his pills for osteoporosis; however, he intentionally dismisses the reminders. So agent `MCIA` has to try to convince him to follow the treatment, he may use one of the following rhetorical arguments:

- ap_{md}: *If you take your medicine, I will talk with your son to come to visit you more frequently and you can talk more with him.*
- th_{md}: *You should take your medicine, otherwise I will tell your son not to bring your preferred cake.*
- rw_{md}: *If you take your medicine, I will talk with the administrator to buy you the rocking chair you want so much.*

Next, we present the mental state of agent `MCIA` and the logical formalization of the rhetorical arguments.

$\texttt{MCIA} = \langle \mathcal{T}, \mathcal{G}, \mathcal{O}pp, \mathcal{GO}, \mathcal{S}_{\mathcal{O}pp}, \mathcal{S}_{\mathcal{GO}}, \mathcal{A}, \mathcal{AO}, \texttt{REP} \rangle$ where:

$\mathcal{T} = \{\mathcal{F}, \mathcal{S}, \mathcal{D}\}$ such that

$\mathcal{S} = \{take(calcium) \rightarrow visit\_of(son, weekly),$

$visit\_of(son, weekly) \rightarrow talk\_more\_with(son),$

$\neg take(calcium) \rightarrow \neg recieve(pref\_cake),$

$take(calcium) \rightarrow have(rocking\_chair),$

$\mathcal{G} = \{g_4\}$ such that $g_4 = get(John, \text{`}take(calcium)\text{'})$

$\mathcal{O}pp = \{Jonh\}$

$\mathcal{GO}_a = \{go_{11}\}$, $\mathcal{GO}_c = \{go_{12}, go_{13}\}$ such that
$go_{11} = have(rocking\_chair)$, $go_{12} = receive(pref\_cake)$, and $go_{13} = talk\_more\_with(son)$

$\mathcal{S}_{\mathcal{O}pp} = \{(John, 0.85, \{go_{11}, go_{12}, go_{13}\})\}$ such that $\texttt{THRES}(John) = 0.85$, $\{go_{11}\} \in RW_{\mathcal{GO}}$, $\{go_{12}\} \in TH_{\mathcal{GO}}$, and $\{go_{13}\} \in AP_{\mathcal{GO}}$

$\mathcal{S}_{\mathcal{GO}} = \{(go_{11}, 0.85), (go_{12}, 0.7), (go_{13}, 0.9)\}$

$\texttt{REP} = 0.9$

[5] Augmented reality.

Table 4. Strength values of the rhetorical arguments of agent MCIA in the patients medication scenario

Goal	IMP(go)	Status	EFF(go)	ST_BASIC(ap)	ST_COMB(ap)	
go_{11}	0.85	Active	0.75	0.8	0.04	rw_{md}
go_{12}	0.7	Chosen	0.75	0.725	0.036	th_{md}
go_{13}	0.9	Chosen	0.25	0.575	0.029	ap_{md}

From this mental state, the following rhetorical arguments can be generated:

$ap_{md} = \langle T_1, g_4, go_{13} \rangle$ where:

$T_1 \cup \text{SECOND}(g_4) = \{(take(calcium), \emptyset),$

$(visit\_of(son, weekly), take(calcium) \rightarrow visit\_of(son, weekly))\}$

$(talk\_more\_with(son), visit\_of(son, weekly) \rightarrow talk\_more\_with(son))$

$th_{md} = \langle T_2, g_4, go_{12} \rangle$ where:

$T_2 \cup \text{SECOND}(g_4) = \{(\neg take(calcium), \emptyset),$

$(\neg recieve(pref\_cake), \neg take(calcium) \rightarrow \neg recieve(pref\_cake))$

$rw_{md} = \langle T_3, g_4, go_{11} \rangle$ where:

$T_3 \cup \text{SECOND}(g_4) = \{(take(calcium), \emptyset),$

$(have(rocking\_chair), take(calcium) \rightarrow have(rocking\_chair))$

First of all, the credibility of MCIA has to be evaluated. We have that MCIA is considered credible by *John* based on REP(MCIA) > THRES(John) (i.e., $0.9 > 0.85$). Thus, we can proceed to calculate the basic strength values of the rhetorical arguments generated by MCIA. Table 4 shows the basic and combined values of the strength; the combined strength is calculated considering that ACCUR_CRED(MCIA, *John*) $= 0.9 - 0.85 = 0.05$.

Thus, we have that reward rw_{md}—whose opponent's goal is $go_{11} = have(rocking\_chair)$—is the strongest rhetorical argument. Even when goal $go_{13} = talk\_more\_with(son)$ is more important, the fact that it is close to be achieved determines that offering the reward can be more persuasive. We can think in the following manner, go_{13} is chosen, so it is likely that *John* and her son have already agreed to talk more and the only thing missing is to set the dates; on the other hand, go_{11} is only active, this means that it is still missing to evaluate the goal and to allocate resources (e.g., money) for buying the chair. Therefore, John will gain more by having a rocking chair bought, which can impact on the persuasive power of agent MCIA.

6 Empirical experiments

In this section, we present three empirical experiments that aim to evaluate our proposal. For this evaluation, we compare our proposal with its closest alternative approach (i.e., Amgoud 2003; Amgoud & Prade, 2004), which is based on the importance of the opponent's goal to determine the strength of a rhetorical argument. The environment is an abstract one involving just two agents. The input for each experiment is a set of rhetorical arguments. The kind of the rhetorical arguments is not relevant for the experiments because these are focused on comparing the strength measurement model. Regarding the technical details, the experiments were implemented in C++ and the values of the importance and the effectiveness were generated randomly in the interval $[0, 1]$ and the set $\{0, 0.25, 0.5, 0.75, 1\}$, respectively. Thus, for each individual negotiation encounter, these values were always different. The values of the basic and the combined strength were calculated from these values. The answers given by the agents were ruled by the strategy defined in the previous section. Finally, the output of the experiments is

mainly the number of negotiation cycles, the number of exchanged arguments, and the number of reached agreements.

In our experiments, a single simulation run involves 1000 separate negotiation encounters between two agents. For all the negotiations, the agents were paired against agents that use the same mechanism of strength calculation. We call 'BBGP-based agents' the agents that use the strength evaluation model proposed in Section 4 and 'IMP-based agents' the agents that use the strength evaluation model based on the importance of the opponent's goal. We performed negotiations where agents generate 10, 25, 50, 100, 250, 500, 750, and 1000 rhetorical arguments. This means that an agent has at most 10, 25, 50, 100, 250, 500, 750, or 1000 arguments to defend his position. We make the experiments with different amounts of arguments in order to analyze the bias of the efficiency of our proposal. For each setting of number of arguments, the simulation was repeated 10 times. This makes a total of 10 000 encounters for each setting. Finally, the experimental variables that were measured are (i) the number of cycles taken to reach agreements, (ii) the number of agreements made, and (iii) the number of arguments (threats, rewards, and appeals) used.

Next, we describe each one of the experiments that will be presented in the following sub-sections:

1. In the first experiment, the credibility of the agents varies in each negotiation encounter. Thus, in some encounters, both agents are credible, in others only one agent is credible, and in others none of the agents is credible.
2. In the second experiment, we focus on the performance of BBGP-based agents. We compare the efficiency of BBGP-based agents that negotiate in fully informed scenarios with the efficiency of BBGP-based agents that negotiate in partially informed scenarios.
3. In the third experiment, we again focus on the performance of BBGP-based agents. Thus, we compare the behavior of the two ways of measuring the strength, namely the basic and the combined strength.

6.1 Experiment 1

In this experiment, we consider that there are some BBGP-based agents that are credible and others that are not. Besides, we use the basic strength calculation. This leads to three possible situations:

1. Both the proponent and the opponent agents are credible. In this case, a negotiation dialogue begins.
2. The proponent agent is credible, whereas the opponent agent is not credible. In this case, any argument used by the opponent will be evaluated by the proponent due to the opponent low credibility. This means that the proponent does not believe that any of his goals can be threatened/rewarded/appealed. On the other hand, the goals the arguments used by the proponent can impact on the goals of the opponent. Thus, we settled that the opponent has to accept to do the required action.
3. The proponent agent is not credible, whereas the opponent agent is credible. In this case, the negotiation does not even begin, because the proponent will never convince the opponent.

Figures 3 and 4 show the behavior of the variables *number of exchanged arguments* and *number of reached agreements*. Recall that for each experiment, we run 1000 negotiation encounters; however, BBGP-based agents only engage in a negotiation when either both are credible or the proponent is credible. We run experiments taking into account different reputation values for the agents and we have noticed that the less the reputation value is the less the number of negotiation encounters is. This is quite rational because low reputation values mean that it is more difficult that agents engage in a negotiation. For the results presented in this experiment, we used a reputation value of 0.8 for both agents and the thresholds are generated randomly in the interval [0, 1] before each negotiation encounter.

The fact that BBGP-based agents do not engage in all the negotiations impacts on the experimental variables. Thus, the number of exchanged arguments is indeed less in negotiation between BBGP-based agents (Figure 3). We could believe that it may impact negatively on the variable number of reached agreements because IMP-based agents participate in all the possible negotiations, that is, 1000 negotiation encounters, while BBGP-based agents only participate in some negotiation encounters. However,

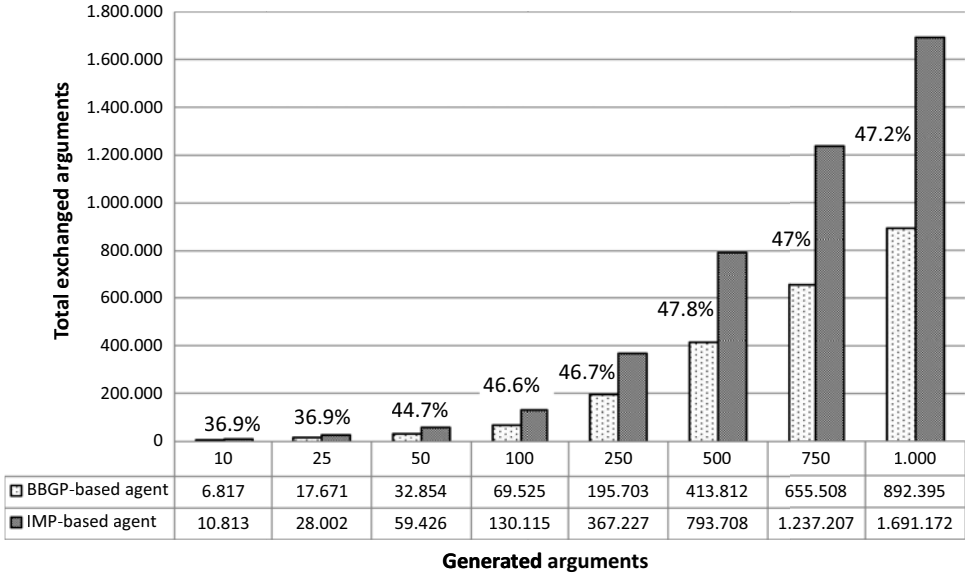

Figure 3 Experiment 1: comparison of the variable *number of exchanged arguments*

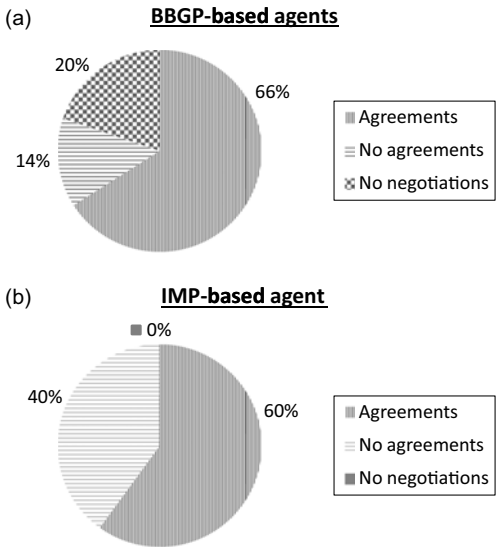

Figure 4 Percentage of negotiations that end in an agreement vs. percentage of negotiations that do not end in an agreement. (a) For BBGP-based agents. (b) For IMP-based agents

the results show that even in that conditions, BBGP-based agents reach more agreements than IMP-based agents. Figure 4 shows the percentages of reached agreements, non-reached agreements, and the percentage of the negotiations the BBGP-agents do not engage in. These values reflect the average behavior of the agents. Notice that, although BBGP-based agents do not engage in all the negotiation, they reach more agreements than IMP-based agents.

6.2 Experiment 2

In this experiment, we focus on comparing the performance of the BBGP-based agents considering that they negotiate either in fully informed scenarios or in partially informed scenarios. Let us recall that in fully informed scenarios the proponent agent employs convincing arguments to try to convince his opponent, whereas in partially informed scenarios the agent does not know which arguments are convincing and which are not.

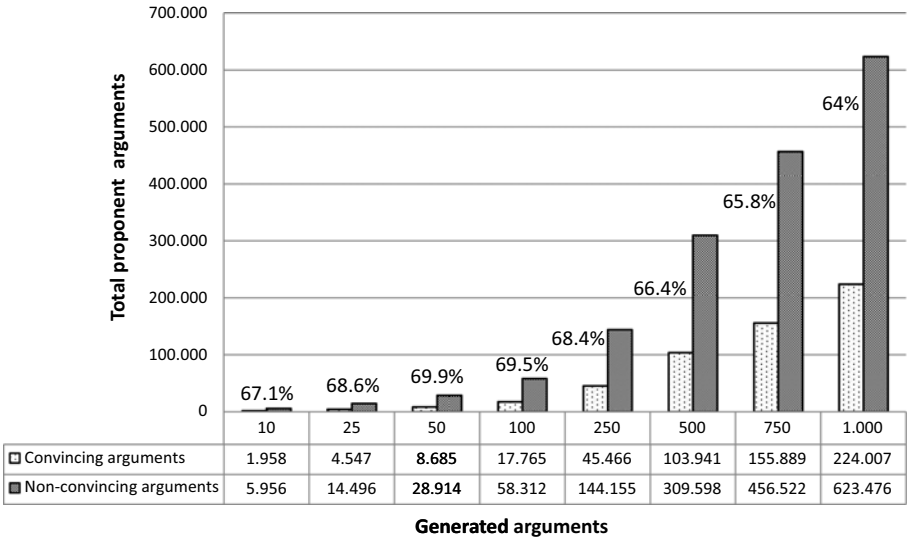

Figure 5 Experiment 2: comparison of the variable *number of arguments sent by the proponent*

The experimental variables that are taken into account in this experiment are the *number of arguments used by the proponent* and the *number of arguments exchanged during the negotiation*. In this experiment, the first variable is especially important given that in fully informed scenarios the proponent knows the value of the required action while in partially informed scenarios the proponent does not know this information. This fact impacts directly on the number of arguments used by the proponent because when he knows this value his persuasive strategy only includes those rhetorical arguments that fulfil the preferability condition, that is, the value of these rhetorical arguments is greater than the value of the action. Thus, agents in fully informed scenarios employ a modified conservative strategy, in which their first rhetorical argument to be sent is the least valued preferable argument.

Figures 5 and 6 show the results of this experiment. In Figure 5, we can observe the behavior of the variable number of proponent arguments. The difference of amount of arguments used by the proponent is very notorious. In average, BBGP-agents in fully informed scenarios use 70 282 arguments, whereas in partially informed scenarios, they use 205 179 arguments during the negotiation encounters. This means that in partially informed scenarios the amount of used arguments is almost three more times than the amount of arguments used in fully informed scenarios. The number of arguments used by the proponent has also an impact on the total number of exchanged arguments (see Figure 6). For this variable, on average, we have that the number of exchanged arguments in fully informed scenarios is 140 030, whereas in partially informed scenarios it is 274 926, which is almost double of arguments.

We defined fully informed scenarios under the premise that the value of the actions is the same for all the participant agents. We are conscientious that in a scenario of robot agents, this premise may be more easily satisfied than in scenarios involving humans. However, with this experiment we wanted to show that the preferability condition has a big impact on the number of arguments used during the negotiation encounters. We have considered that in partially informed scenarios, the proponent agents do not know the value of the actions for their opponents; nevertheless, we could consider that a proponent agent may employ the value he gives to his actions as a reference point to select their arguments. All this aiming at distinguishing convincing arguments and non-convincing arguments.

6.3 Experiment 3

In this experiment, we again evaluate the performance of BBGP-based agents. However, in this experiment, we compare the performance of the basic strength to the performance of the combined strength. Let us recall that for calculating the basic strength, we only take into account the opponent's goal, whereas for calculating the combined strength, besides the opponents goal, we take into account the accurate

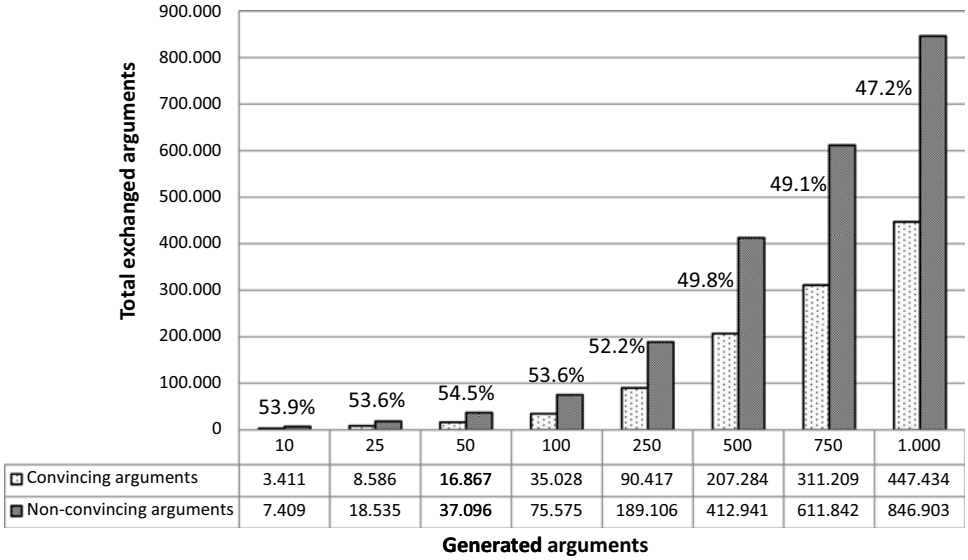

Figure 6 Experiment 2: comparison of the variable *number of exchanged arguments*

credibility of the agent. Recall that the reputation is an evidence of the proponent's past behavior of an agent with respect to his opponents. We assume that this value is already estimated and it is not private information; thus, the reputation value of an agent is visible for any other agent. On the other hand, the 'accurate' value of the credibility of an agent P with respect to an opponent O—whose threshold is $\texttt{THRES}(O)$—is given by $\texttt{ACCUR\_CRED}(P, O) = \texttt{REP}(P) - \texttt{THRES}(O)$. Thus, we want to know how the value of the accurate credibility impacts on the studied variables, that is, the number of arguments sent by the proponent and the number of exchanged arguments during the dialogue.

The settings in this experiment are a little different from the settings in previous experiments. This is because we want to focus on the impact of the accurate credibility. Thus, the settings of this experiment are each agent generates 10 arguments, the reputation of each agents is 1, the threshold of the proponent agent is always 0.8, and the threshold of the opponent goes down from 0.8 to 0.55. Decreasing the opponent's threshold makes him more credulous and persuadable. We have run six scenarios:

- In the first scenario, both agents have the same reputation and threshold values; therefore, the value of the accurate credibility is the same for both agents (i.e., $1 - 0.8 = 0.2$).
- In the second scenario, the threshold of the proponent agent is 0.8 and the threshold of the opponent is 0.75; hence, the value of the accurate credibility is different. Thus, the value of the accurate credibility that the proponent uses for calculating the combined strength of the arguments is 0.25, whereas the value of the accurate credibility that the opponent uses for calculating the combined strength of the arguments is 0.2. This means that there is a difference of 0.05 between both values of the accurate credibility and this also means that in the second scenario the opponent is more persuadable than in the first scenario.
- In the remaining scenarios, the threshold of the proponent is 0.8 and the threshold of the proponent goes down in 0.05 in each scenario. Thus, in the third scenario, the threshold of the opponent is 0.7, in the fourth scenario it is 0.65, in the fifth scenario it is 0.6, and in the sixth scenario it is 0.55. This means that the difference between the value of the accurate credibility increases. Thus, in the third scenario, the difference is 0.1, in the fourth scenario it is 0.15, in the fifth scenario it is 0.2, and in the last scenario it is 0.25.

Table 5 shows a resume of these six scenarios, which includes the values of the reputation, threshold, and accurate credibility of agents proponent and opponent.

We have run 1000 negotiation encounters for each scenario and besides evaluating the variables of efficiency; we also compare the number of times that the proponent succeeds in persuading the opponent.

Table 5. Experiment 3: values of the reputation, threshold, and accurate credibility of agents proponent and opponent

	Proponent			Opponent		
Scenario	REP	THRES	ACCUR_CRED	REP	THRES	ACCUR_CRED
#1	1	0.8	0.2	1	0.8	0.2
#2	1	0.8	0.2	1	0.75	0.25
#3	1	0.8	0.2	1	0.7	0.3
#4	1	0.8	0.2	1	0.65	0.35
#5	1	0.8	0.2	1	0.6	0.4
#6	1	0.8	0.2	1	0.55	0.45

Table 6. Percentage of negotiations that end favorably for the proponent *P* vs. percentage of negotiations that end favorably for the opponent *O*

	Basic strength		Combined strength	
$\lvert ACCUR_O - ACCUR_P \rvert$	O	P	O	P
0	49%	51%	49%	51%
0.05	49%	51%	10%	90%
0.1	49%	51%	4.56%	95.44%
0.15	49%	51%	0.32%	99.68%
0.2	49.6%	50.4%	0.002%	99.998%
0.25	49.5%	50.5%	0%	100%

Figures 7 and 8 illustrate the results of this experiment. Figure 7 shows the behavior of the variable number of arguments sent by the proponent. When the calculation is done by applying the combined strength equation, we can notice that the greater the difference between the values of the accurate credibility, the fewer the number of arguments the proponent sends. On the other hand, when the calculation is done by applying the basic strength equation, the number of arguments is very similar. Figure 8 shows the behavior of the variable number of exchanged arguments. This result reaffirms that the performance of the combined strength calculation improves as the difference of the values of the accurate credibility increases.

In Table 6, we compare the number of times that the proponent succeeds in persuading the opponent. Each line corresponds to one of the scenarios defined in Table 5. In all the scenarios, when the calculation is based on the basic strength, the percentages of succeed of proponent and opponent are balanced whereas when the negotiation is on the combine strength, the percentage of succeed of the proponent increases as the difference of the accurate credibility values increases. The difference in success percentage is even more notorious in the last scenarios. Indeed, in the fourth scenario the percentage of succeed of the proponent is almost 100% and in the last scenario it is 100%. We can say that if the difference between the values of the accurate credibility is equal to or greater than 0.2, the proponent always succeeds.

We can conclude that while it is true that our approach has a better performance, it is also true that it is necessary to model further knowledge about the opponent. This need of further modeling may be seen as a weakness of the model; however, we can notice that in all the evaluated variables, the model is always more efficient and effective than the other approach. Indeed, when more criteria are employed both the efficiency and the effectiveness increase. Let us recall Experiment 1, IMP-based agents engage in all negotiation encounters (i.e., 1000 encounters) whereas BBGP-based agents only engage in a negotiation

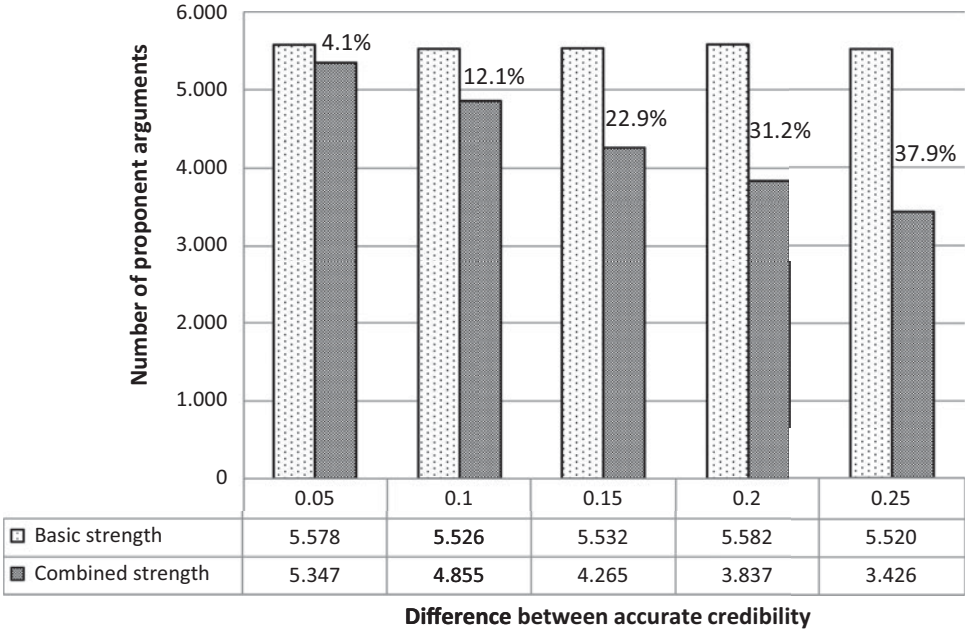

Figure 7 Experiment 3: comparison of the variable *number of arguments sent by the proponent*

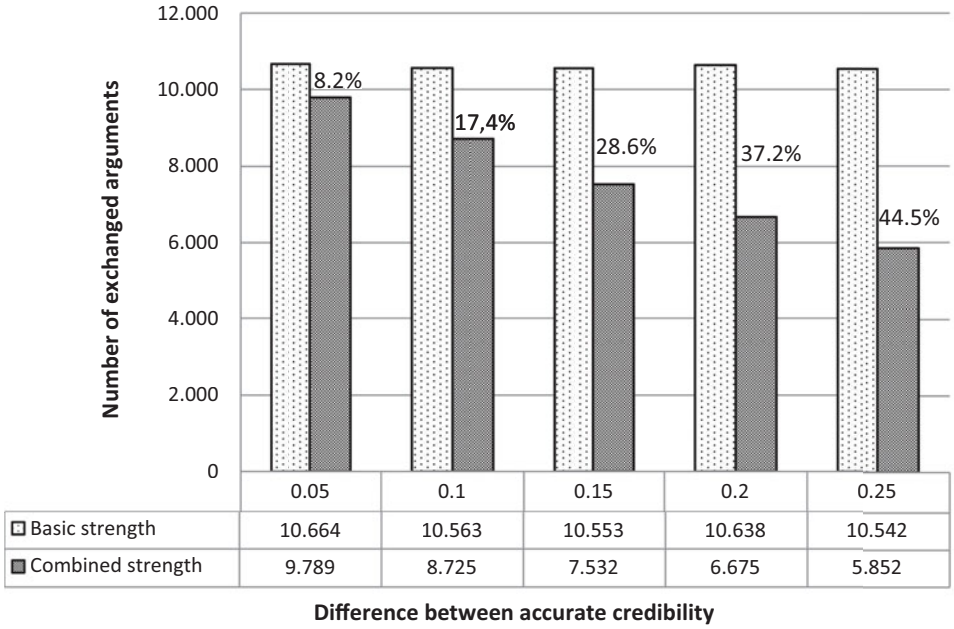

Figure 8 Experiment 3: comparison of the variable *number of exchanged arguments*

encounter when the proponent agent is credible enough (i.e., around 800 encounters). This would mean that IMP-based agents may always achieve more agreements than BBGP-based agents; nevertheless, the results show that the higher the number of generated arguments is, the more agreements the BBGP-agents achieve. The efficiency of the model is even more notorious when the scenario is a fully informed scenario. In this type of scenario, the number of negotiation cycles and the number of exchanged arguments are less than in partial informed scenarios. Thus, when we take into account the criterion credibility and the type of scenario, the efficiency of the model increases even more.

7 Discussion

This section presents the main related work and the differences with our proposal. Besides, considering that persuasion has a human–computer interaction aspect, we discuss how researchers in this community measure persuasion.

The most related work is the research made by Amgoud and Prade (2004, 2005b, 2006). They propose a formal definition of rhetorical arguments and a strength evaluation system. For the evaluation of the strength of rhetorical argument, the certainty of the beliefs that are used for the generation of the argument and the importance of the opponents goal are considered. In our proposal, we further analyze the components of a rhetorical argument and suggested new criteria for calculating the strength values. We also proposed a set of steps to be considered for the calculation of the strength values. Both the criteria and the steps are inspired on the work of Guerini and Castelfranchi (2006).

Another related work is presented by Ramchurn *et al*. (2003). They propose a model where the strength value of rhetorical arguments varies during the negotiation depending on the environmental conditions. For calculating the strength value of the argument, it is taken into account a set of world states an agent can be carried to by using a certain argument. The intensity of the strength values depends on the desirability of each of these states. For a fair calculation, an average over all possible states is used. The difference with our proposal is that their proposal is not based on logical language and they do not consider the components of the arguments.

We have also worked on this topic. In a preliminary work, we focused on calculating the strength value of threats. In Morveli-Espinoza *et al*. (2016), we proposed a way for calculating the basic strength of threats. An extended article—Morveli-Espinoza *et al*. (2019)—considered the status of the opponent goal and the credibility of the opponent; however, it was not taken into account the preferability (which makes difference between rewards/appeals and threats) and the difference between fully and partially informed scenarios. Finally, Morveli-Espinoza *et al*. (2020) is a previous version of this article whose differences were detailed in Introduction.

When persuasion involves humans, arguments have the form of messages, which may be (or not) threats, rewards, or appeals. Aside of the form of the messages, it is interesting to know the way persuasiveness is measured in this context, specially because we plan to extend our research by involving humans. In the literature, we can notice that two kinds of persuasiveness are measured, namely the actual persuasiveness and the perceived persuasiveness. For the former, it is measured if message produced the intended persuasive effects on attitudes, intentions, or behaviors, whereas for the latter, the people's perceptions about the influence of message on them are measured (OKeefe 2018).

Most of the work in the literature uses scales to measure the persuasiveness of messages. Thomas *et al*. (2019) present a list of messages and scales to evaluate them, which was collected from several articles. It can be distinguished that some articles measure the susceptibility of participants to some Cialdini's principles[6] (e.g., Kaptein *et al*. 2009), others the persuasive power of messages (e.g., Thomas *et al*. 2019; Allen *et al*. 2000), and others measure the persuasibility of persuasive strategies (e.g., Busch *et al*. 2013). Thus, we can notice that the persuasion is measured by considering only the message or the impact of it on the participants. In our case, we measure the message by evaluating its components, which can have an impact on the opponent (participant).

8 Conclusions and future work

In this work, we proposed a model for the strength calculation of rhetorical arguments. We studied the pre-conditions for an argument to be considered convincing. We based on the proposal of Guerini and Castelfranchi (2006), who claimed that the credibility of the proponent and the preferability of the opponent's goal over the value of the required actions determine convincing arguments. We used the reputation of the proponent and the threshold of trust of the opponent to evaluate the credibility of the

[6] Cialdini (2007) claims that the influence is based on six key principles: reciprocity, commitment and consistency, social proof, authority, liking, and scarcity. Furthermore, a seventh principle—called the unity principle—was added (Cialdini 2016).

proponent and the opponent's goal importance and its status to evaluate the preferability. We did not directly use the status of an opponent's goal, but we judge its effectiveness based on the type of rhetorical argument and the status itself. Based on the numerical values of these pre-conditions, we have proposed a model for evaluating and calculating the strength value of the rhetorical arguments. The model starts evaluating the credibility of the proponent agent. The proponent agent can proceed to the calculation of the rhetorical arguments only if he is considered credible by his opponent; otherwise, the process ends.

A set of experiments were done with the aim to evaluate *the number of exchanged arguments* and *the number of reached agreements*. In all cases, we demonstrated that our proposed model fares better than the calculation model that only takes into account the importance of the opponent's goal. Furthermore, we evaluated the credibility of the agents before the calculation task begins. Thus, we have noticed that the new criteria included in the calculation model have made our proposal more efficient than the model based only on one criteria.

We worked under the premise that the proponent agent knows in advance the information about his opponent. An interesting future work is to complement this model with the study of an adequate opponent modeling approach. We can also consider that the information of the opponent is uncertain, which may impact on directly the strength calculation. In the proposed approach, there is no model of the environment or the context where the negotiation occurs, especially in terms of organizational structure. We believe that this information can influence the strength of the arguments and therefore on the persuasion power of the agents.

Finally, as said before, we plan to make experiments with humans; more concretely, we plan to make experiments in the patients medication scenario by using HoloLens for the interaction with the humans since the use of this technology is comfortable for adults younger than 65 years, as was demonstrated by Blusi & Nieves (2019).

Acknowledgements

This work was partially financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Brazil. Juan Carlos Nieves was partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement n° 825619 (AI4EU project).

References

Allen, M., Bruflat, R., Fucilla, R., Kramer, M., McKellips, S., Ryan, D. J. & Spiegelhoff, M. 2000. Testing the persuasiveness of evidence: combining narrative and statistical forms. *Communication Research Reports* **17**(4), 331–336.

Amgoud, L. 2003. A formal framework for handling conflicting desires. In *ECSQARU*, **2711**, 552–563. Springer.

Amgoud, L. & Besnard, P. 2013. A formal characterization of the outcomes of rule-based argumentation systems. In *International Conference on Scalable Uncertainty Management*, 78–91. Springer.

Amgoud, L., Parsons, S. & Maudet, N. 2000. Arguments, dialogue, and negotiation. In *Proceedings of the 14th European Conference on Artificial Intelligence*, 338–342.

Amgoud, L. & Prade, H. 2004. Threat, reward and explanatory arguments: generation and evaluation. In *Proceedings of the ECAI Workshop on Computational Models of Natural Argument*, 73–76.

Amgoud, L. & Prade, H. 2005a. Formal handling of threats and rewards in a negotiation dialogue. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 529–536. ACM.

Amgoud, L. & Prade, H. 2005b. Handling threats, rewards, and explanatory arguments in a unified setting. *International Journal of Intelligent Systems* **20**(12), 1195–1218.

Amgoud, L. & Prade, H. 2006. Formal handling of threats and rewards in a negotiation dialogue. In *Argumentation in Multi-Agent Systems*, 88–103. Springer.

Baarslag, T., Hendrikx, M. J., Hindriks, K. V. & Jonker, C. M. 2016. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques. *Autonomous Agents and Multi-Agent Systems* **30**(5), 849–898.

Blusi, M. & Nieves, J. C. 2019. Feasibility and acceptability of smart augmented reality assisting patients with medication pillbox self-management. *In Studies in Health Technology and Informatics*, 521–525.

Busch, M., Schrammel, J. & Tscheligi, M. 2013. Personalized persuasive technology–development and validation of scales for measuring persuadability. In *International Conference on Persuasive Technology*, 33–38. Springer.

Castelfranchi, C. & Guerini, M. 2007. Is it a promise or a threat? *Pragmatics & Cognition* **15**(2), 277–311.

Castelfranchi, C. & Paglieri, F. 2007. The role of beliefs in goal dynamics: prolegomena to a constructive theory of intentions. *Synthese* **155**(2), 237–263.

Cialdini, R. 2016. *Pre-Suasion: A Revolutionary Way to Influence and Persuade*. Simon and Schuster.

Cialdini, R. B. 2007. *Influence: The psychology of persuasion*, **55**. Collins.

Dimopoulos, Y. & Moraitis, P. 2011. Advances in argumentation based negotiation. In *Negotiation and Argumentation in Multi-agent Systems: Fundamentals, Theories, Systems and Applications*, 82–125.

Falcone, R. & Castelfranchi, C. 2001. Social trust: a cognitive approach. In *Trust and Deception in Virtual Societies*, 55–90. Springer.

Falcone, R. & Castelfranchi, C. 2004. Trust dynamics: how trust is influenced by direct experiences and by trust itself. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, 740–747. IEEE.

Florea, A. M. & Kalisz, E. 2007. Adaptive negotiation based on rewards and regret in a multi-agent environment. In *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 2007. SYNASC*, 254–259. IEEE.

Guerini, M. & Castelfranchi, C. 2006. Promises and threats in persuasion. In *6th Workshop on Computational Models of Natural Argument*, 14–21.

Hadjinikolis, C., Modgil, S. & Black, E. 2015. Building support-based opponent models in persuasion dialogues. In *International Workshop on Theories and Applications of Formal Argumentation*, 128–145. Springer.

Hadjinikolis, C., Siantos, Y., Modgil, S., Black, E. & McBurney, P. 2013. Opponent modelling in persuasion dialogues. In *IJCAI*.

Hunter, A. 2015. Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 3055–3061.

Ingeson, M., Blusi, M. & Nieves, J. C. 2018. Microsoft hololens-a mhealth solution for medication adherence. In *International Workshop on Artificial Intelligence in Health*, 99–115. Springer.

Kaptein, M., Markopoulos, P., de Ruyter, B. & Aarts, E. 2009. Can you be persuaded? individual differences in susceptibility to persuasion. In *IFIP Conference on Human-Computer Interaction*, 115–118. Springer.

Lam, H.-P. & Governatori, G. 2011. What are the necessity rules in defeasible reasoning? In *International Conference on Logic Programming and Nonmonotonic Reasoning*, 187–192. Springer.

Morveli-Espinoza, M., Nieves, J. C. & Tacla, C. A. 2020. Measuring the strength of rhetorical arguments. In *To be published in the Proceedings of the 17th European Conference on Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems.

Morveli-Espinoza, M., Possebom, A. T. & Tacla, C. A. 2016. Construction and strength calculation of threats. In *Computational Models of Argument - Proceedings of COMMA 2016, Potsdam, Germany, 12–16 September,* 2016, 403–410.

Morveli-Espinoza, M., Possebom, A. T. & Tacla, C. A. 2019. On the calculation of the strength of threats. *Knowledge and Information Systems* **62**(4), 1511–1538.

OKeefe, D. J. 2018. Message pretesting using assessments of expected or perceived persuasiveness: evidence about diagnosticity of relative actual persuasiveness. *Journal of Communication* **68**(1), 120–142.

Pinyol, I. & Sabater-Mir, J. 2013. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review* **40**(1), 1–25.

Rahwan, I., Ramchurn, S. D., Jennings, N. R., Mcburney, P., Parsons, S. & Sonenberg, L. 2003. Argumentation-based negotiation. *The Knowledge Engineering Review* **18**(04), 343–375.

Ramchurn, S. D., Jennings, N. R. & Sierra, C. 2003. Persuasive negotiation for autonomous agents: a rhetorical approach. In *Proceedings of the Workshop on Computational Models of Natural Argument*, 9–17.

Ramchurn, S. D., Sierra, C., Godo, L. & Jennings, N. R. 2007. Negotiating using rewards. *Artificial Intelligence* **171**(10–15), 805–837.

Rienstra, T., Thimm, M. & Oren, N. 2013. Opponent models with uncertainty for strategic argumentation. In *IJCAI*.

Sabater, J. & Sierra, C. 2001. Regret: a reputation model for gregarious societies. In *Proceedings of the 4th Workshop on Deception Fraud and Trust in Agent Societies*, 70, 61–69.

Shi, B., Tao, X. & Lu, J. 2006. Rewards-based negotiation for providing context information. In *Proceedings of the 4th International Workshop on Middleware for Pervasive and Ad-Hoc Computing*, 8. ACM.

Sierra, C., Jennings, N. R., Noriega, P. & Parsons, S. 1997. A framework for argumentation-based negotiation. In *International Workshop on Agent Theories, Architectures, and Languages*, 177–192. Springer.

Sierra, C., Jennings, N. R., Noriega, P. & Parsons, S. 1998. A framework for argumentation-based negotiation. In *Intelligent Agents IV Agent Theories, Architectures, and Languages*, 177–192. Springer.

Sycara, K. P. 1990. Persuasive argumentation in negotiation. *Theory and Decision* **28**(3), 203–242.

Thomas, R. J., Masthoff, J. & Oren, N. 2019. Can i influence you? development of a scale to measure perceived persuasiveness and two studies showing the use of the scale. *Frontiers in Artificial Intelligence* **2**, 24.

Yu, B. & Singh, M. P. 2000. A social mechanism of reputation management in electronic communities. In *International Workshop on Cooperative Information Agents*, 154–165. Springer.