International Journal of Approximate Reasoning 54 (2013) 667–689

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Arguing about social evaluations: From theory to experimentation

Isaac Pinyol [a,b,*], Jordi Sabater-Mir [c]

[a] iMathResearch S.L., Centre de Recerca Matematica, Campus UAB, Bellaterra, Barcelona, Spain
[b] ASCAMM Technology Center, Av. Universitat Autonoma, 23 (08290) Cerdanyola del Valles, Barcelona, Spain
[c] IIIA – CSIC, Artificial Intelligence Research Institute, Spanish National Research Council, Campus UAB, Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:
Received 29 July 2011
Received in revised form 17 November 2012
Accepted 23 November 2012
Available online 8 December 2012

Keywords:
Multi-agent systems
Reputation
Argumentation-based protocol
Trust
Reliability measures

ABSTRACT

In open multiagent systems, agents depend on reputation and trust mechanisms to evaluate the behavior of potential partners. Often these evaluations are associated with a measure of reliability that the source agent computes. However, due to the subjectivity of reputation-related information, this can lead to serious problems when considering communicated social evaluations. In this paper, instead of considering only reliability measures computed from the sources, we provide a mechanism that allows the recipient decide whether the piece of information is reliable according to its own knowledge. We do it by allowing the agents engage in an argumentation-based dialog specifically designed for the exchange of social evaluations. We evaluate our framework through simulations. The results show that in most of the checked conditions, agents that use our dialog framework significantly improve (statistically) the accuracy of the evaluations, over the agents that do not use it. In particular, the simulations reveal that when there is a heterogeneity set of agents (not all the agents have the same goals) and agents base part of their inferences on third-party information, it is worth using our dialog protocol.

1. Introduction and motivation

Computational trust and reputation models have been recognized as one of the key technologies in the design of open multi-agent systems [1]. These models provide social evaluations about the potential performance of agents in a specific context, by aggregating (mainly) the outcomes of past interactions and third-party communications. Some models also attach to the social evaluation a reliability measure that reflects how *confident* the source agent feels about that value. This allows agents to internally weight the relevance of the calculated evaluations.

The reliability value is transmitted together with the social evaluation when there is a communication, so the recipient agent can decide whether it is worth considering that piece of information. However, due to the subjectivity of reputation information, a social evaluation declared reliable by agent *A* may not be reliable for agent *B*, because the bases under which *A* has inferred the evaluation cannot be accepted by *B*. This can happen because agents have different mechanisms to infer social evaluations, have had different experiences, have different goals, etc. The use of reliability measures in communicated social evaluations is restricted to those situations where the recipient agent knows that the source agent is honest, that has a similar way to calculate social evaluations in that specific context and that has had similar experiences.

This paper offers an alternative mechanism. We suggest that the reliability measure cannot depend on the source agent, but must be fully evaluated by the recipient agent according to its own knowledge. In our approach, rather than only allow one shot communications, we allow agents to participate in argumentation-based dialogs regarding reputation elements

* Corresponding author at: ASCAMM Technology Center, Av. Universitat Autonoma, 23 (08290) Cerdanyola del Valles, Barcelona, Spain.
E-mail addresses: ipinyol@imathresearch.com, ipinyol@ascamm.com (I. Pinyol), jsabater@iiia.csic.es (J. Sabater-Mir).

0888-613X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2012.11.006

in order to decide on the reliability (and thus acceptance) of a communicated social evaluation. Our approach differs from others in that it is the recipient agent, not the source agent, who decides about the reliability of a communicated evaluation. [1]

Although we assume that agents use different reputation models, we consider that they use a common language to express reputation-related concepts. Such a language is defined in Section 2. We introduce in the same section the concepts of *reputation theory* to characterize the structure of reputation-related information that we handle in this paper. In section 3, we exemplify the class of problems that we solve in this paper and give the main features of the proposed framework. Since our approach uses argumentation techniques we define the notion of argument and attack between arguments regarding reputation-related information in Section 4. We also provide a method for deciding whether a communicated social evaluation can be considered reliable enough by the recipient agent. We specify completely the dialog protocol in Section 5, and in Section 6 we demonstrate with experimentation that the proposed mechanism significantly improves the accuracy of the evaluations. Section 7 details the related work and Section 8 concludes the analysis and introduces work for the future.

2. The L_{rep} language

L_{rep} is a language that captures the reputation-related information that individual agents use to write statements (and reason) about reputation concepts. The language is based on an ontology of reputation [3] used to characterized the reputation information.

The main element we are interested in representing is a *social evaluation*. From a cognitive perspective a social evaluation is a belief that encodes an evaluation of a social entity in a given context [4,5]. In a more computational fashion and according to [3,4], social evaluations incorporate three main elements: the target, the context, and the value of the evaluation. For instance, a social evaluation may say that an agent a (target), as a car driver (context) is very good (value).

From the concept of *social evaluation*, a taxonomy of social evaluations is defined, including for instance, the concepts of *image* and *reputation*, that we are interested to capture within the L_{rep} language. The following subsections formally describe the language as a many-sorted first-order language, giving a brief description of each type of social evaluation.

2.1. Defining L_{rep}

Following [6] where languages are built as a hierarchy of first-order languages, we define $L_{context}$, and L_{rep}. Both are classical first-order languages with equality and contain the logical symbols $\wedge$, $\neg$ and $\rightarrow$. [2] $L_{context}$ is the language that the agents use to describe the context of the evaluations, like norms, or skills, while L_{rep} is used to write statements about reputation.

Definition ($L_{context}$-Domain Language)**.** $L_{context}$ is an unsorted first-order language that includes predicates, constants and functions, necessary for writing statements about the domain. Even when we do not provide any specific language for describing the context of the evaluations, we suggest that a first-order language should be enough to express norms, standards or skills.

Definition (L_{rep}-Reputation language)**.** L_{rep} is a sorted first-order language that agents use to reason about social evaluations. It includes $L_{context}$ and special first-order predicates that are identified by their sorts. These special predicates describe the types of social evaluations, (*Image, Reputation, Shared Voice, Shared Evaluation, Direct Experience*) and Communications (Img, Rep, ShV, ShE, DE and Comm from now on). From now on, direct experiences and communications will be called *ground elements*, the basic elements from which social evaluations are inferred.

The taxonomy of social evaluations appears for the first time in [4], but is formalized in [3]. In Sections 2.3 and 2.4 we present some examples on how L_{rep} can be used to model the information managed by the eBay model [7] and Abdul-Rahman and Hailes model [8] respectively.

The sorts that the language uses are the following:

- S_A: It includes a finite set of target identifiers $\{i_1, \ldots, i_n\}$, which embraces single agents, group of agents and institutions. In fact, we assume that each possible group has assigned an identifier.
- S_F: It contains the set of constant formulas representing elements of $L_{context}$ and L_{rep} itself. The idea is that well-formed formulas from $L_{context}$ and L_{rep} are introduced in L_{rep} as constants for the language. [3] In this way, they can be nested in a first-order predicate. Regarding embedded L_{rep} formulas we only allow one nested level. We use it to capture the idea of communicated social evaluations.
- S_V: It represents the values of the evaluation. Our needs require that the set of possible values is countable, and that a linear pre-order is defined between the values.

[1] A preliminary version of the theoretical development was published in [2].

[2] For the sake of clarity we reduce the first-order languages to facts, conjunctions of facts, and rules.

[3] It can be built recursively and simultaneously with S_F. We add the constant $\ulcorner\varphi\urcorner$ for each $\varphi \in wff(L_{context})$ and the constant $\ulcorner\Psi\urcorner$ for each formula $\Psi \in wff(L_{rep})$.

- S_T: It incorporates discrete time instants. We use them to express that direct experiences and communications take place in a *discrete* unit of time. In a more pragmatic view, it also serves as a unique identifier for the communication and direct interactions.

We pay special attention to the sort S_V, which represents values of a totally ordered set $M = \langle G, \leq \rangle$. It includes the set of constants C_V containing a label v for each $v \in G$. Examples of M are $\langle [0, 1] \cap \mathbb{Q}, \leq \rangle$, where $\leq$ is the standard pre-order binary function for rational numbers, and $\langle \{VB, B, N, G, VG\}, \leq_s \rangle$ referring to the linguistic labels *Very Bad, Bad, Neutral, Good, Very Good*, where $VB \leq_s B \leq_s N \leq_s G \leq_s VG$.

The set of well-formed formulas of L_{rep} (wff(L_{rep})) is defined using the standard syntax of classical first-order logic, and includes special predicates to describe reputation-related concepts. They are *Img*, *Rep*, *ShV*, *ShE*, *DE*, *Comm*. As mentioned before, the last two predicates (*DE* and *Comm*) are what we call *ground elements*.

- $Img(S_A, S_F, S_V)$: It represents an image predicate, an evaluation that is believed by an agent. For instance,

$$Img(j, \ulcorner Provider(service(X)) \urcorner, VG)$$

  indicates that the agent holding the predicate has a *VG* image of agent *j* as a provider of service X. In terms of the mental state of the agent, it indicates that and the holder of the predicate *believes* such evaluation. In this case and in future examples, we take M as $< VB, B, N, G, VG, \leq_s >$ where the elements represent linguistic labels indicating *very bad*, *bad*, *neutral*, *good* and *very good*.
- $Rep(S_A, S_F, S_V)$: It represents a reputation predicate. A reputation refers to an evaluation that is known to circulate in the society. For example

$$Rep(j, \ulcorner Provider(service(X)) \urcorner, VG)$$

  indicates that "it is said" that agent *j* is *VG* as a provider of service X. In this case, the agent holding the predicate believes that the evaluation circulates in society, but this does not imply that the agent believes the evaluation.
- $ShV(S_A, S_F, S_V, S_A)$, $ShE(S_A, S_F, S_V, S_A)$: They represents a shared voice and a shared image respectively. A shared voice is also an evaluation that circulates in the society (like reputation). The difference is that the members of the society that *say* it, are identified. A shared image is a belief about the beliefs of other agents. It indicates that the holder of the predicate believes that a certain group of identified agents *believe* an evaluation. Both predicates include the group that shares the voice or image respectively.
- $DE(S_A, S_F, S_V, S_T)$: It represents a direct experience. For instance,

$$DE(j, \ulcorner Provider(service(X)) \urcorner, VG, t_2)$$

  indicates that the agent had a *VG* direct experience with *j* as a Provider of service X at the time t_2.
- $Comm(S_A, S_F, S_T)$: It represents a communication. For example,

$$Comm(j, \ulcorner Img(j, k, Provider(service(X)), VG) \urcorner, t_2)$$

  indicates that the agent received a communication at time t_2 from agent *j* saying that its image about *k* as a Provider of service *X* is *VG*.

Often, we will write a subindex to explicitly state the agent holding the predicate. For instance,

$$DE_i(j, \ulcorner Provider(service(X)) \urcorner, VG, t_2)$$

indicates that agent *i* has had a direct experience with *j* as a provider of service X at the time t_2 and it was *VG*.

2.2. Reputation theories

We define the concept of reputation theory to characterize all the reputation information that an agent *i* holds. Intuitively, we consider that from a set of direct experiences (DE) and communications (Comm) (the ground elements) agents are able to infer the remaining reputation information (image, reputation, shared voice and shared evaluation) through a consequence relation $\vdash_i$, associated with agent *i*. The consequence relation can be understood as the agent *i*'s reputation model. Formally:

Definition (Reputation theory)**.** Let $\Delta \subset wff(L_{rep})$, we say that Δ is a reputation theory when $\forall \alpha \in \Delta$, α is a ground element. Then, letting $d \in wff(L_{rep})$, we write $\Delta \vdash d$ to indicate that from the reputation theory Δ, d can be deduced via $\vdash$.

The reputation-related information that agent i holds is characterized by the tuple

$$\langle \Delta_i, \vdash_i \rangle$$

where Δ_i is i's reputation theory, and $\vdash_i$ the consequence relation (i's reputation model). The semantics of the language is given by the reputation model, and the axiomatization by the consequence relation $\vdash$ associated with a reputation model. The importance for us relies in the capability of the language to capture the information that reputation models manage.

2.3. Example 1: eBay reputation model

eBay site [7] is one of the most popular on-line marketplace, with more than 100 million registered users. The site offers several services that enhance the usability of the marketplace, like secured pay-pal, on-line auctions, categorical searches etc. Also, it offers a reputation service: it allows the buyers to rate the seller once a transaction is finished, and summarizes this information publicly, for other potential buyers. The eBay reputation model considers reputation as a public and centralized value in which the context is implicit. Buyers rate sellers after each transaction with values of $+1, 0, -1$. The reputation value of each seller is calculated as the sum of all the ratings over the last six months, and presented to potential buyers with a system of colored stars. For instance, it uses a golden star for top sellers and a purple star for sellers with a score between 500 and 999.

Considering now L_{rep}, the reputation theory in eBay's model is composed of a set of communicated direct experiences, where the ratings from buyers are the direct experiences, and the context is constant (say C), and the value representation is a bounded rational type ($[0, 1] \cap \mathbb{Q}$). We can easily normalize the values $-1, 0, 1$ to $0, 0.5, 1$ respectively. As an example, let $b_1, b_2, \ldots$ be buyers, and $s_1, s_2, \ldots$ sellers, a reputation theory for the eBay system could then have the following elements:

$$Comm(b_1, DE(s_1, C, 0, t_1))$$
$$Comm(b_2, DE(s_1, C, 0, t_2))$$
$$Comm(b_2, DE(s_1, C, 0.5, t_3))$$

$$Comm(b_1, DE(s_2, C, 1, t_4))$$
$$Comm(b_4, DE(s_2, C, 1, t_5))$$
$$Comm(b_3, DE(s_2, C, 1, t_6))$$

The model computes the reputation of each sellers. Since eBay score goes from 0 to 100,000, a simple normalized transformation to the interval [0,1] seems plausible. However, notice that the colored stars representation does not follow a linear curve. From a semantic point of view and in our value representation, 0 means very bad reputation, 0.5 neutral reputation, and 1 very good reputation. Having more than 10 points is already considered a good reputation in eBay's model. The next step in the scale is more than 100 points (with a different colored star), and the next is more than 500. In conclusion, there is no linear relation between the punctuation and the semantic representation of the stars. A possible transformation function is described in the following equation:

$$H : [0, 100000] \to [0, 1] \tag{1}$$

$$H(X) = \begin{cases} 0 & \text{if } X < 10; \\ 1 & \text{if } X > 100000; \\ \frac{log(X)-0.5}{8} + 0.5 & \text{otherwise.} \end{cases} \tag{2}$$

The idea is that from a set of communicated direct experiences, reputation predicates can be inferred through $\vdash_{eBay}$. According to the previous reputation theory example, the generated predicates are

$$Rep(s_1, C, 0)$$
$$Rep(s_2, C, 0)$$

In the example, s_2 gets a punctuation of 0 because its punctuation is still lower than 10.

2.4. Example 2: The Abdul-Rahman and Hailes model

The model presented by Abdul-Rahman and Hailes [8] uses the term *trust*, and evaluations take into account the context. The model is fed by two sources: direct experiences and third-party communications of direct experiences. The representation of the evaluations is done in terms of the discrete set *{vt (very trustworthy), t (trustworthy), u (untrustworthy), vu (very untrustworthy)}*. For each agent and context the system keeps a tuple with the number of past own experiences or communicated experiences in each category. For instance, agent A may have a tuple of agent B as a seller like (0, 0, 2, 3),

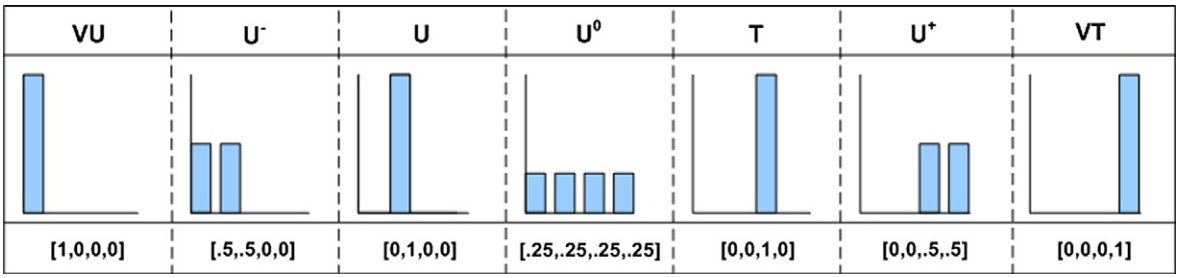

Fig. 1. Abdul-Rahman and Hailes model values expressed in terms of a probabilistic distribution.

meaning that agent *A* has received or experienced 2 results as untrustworthiness and 3 as very untrustworthiness. Finally the *trust* value is computed taking the maximum of the tuple values. In our example, agent B as a seller would be very untrustworthy according to A.

In the case of tie between *vt* and *t* and between *u* and *vu* the system gives the values U^+ (mostly trustworthy) and U^- (mostly untrustworthy) respectively. In any other tie case the system returns U^0 (neutral).

Each agent holds its own reputation theory. In this case, we use a specific representation type for the evaluation values of the model. Similar to the Repage model [9], we define a probability distribution over the ordered set of linguistic labels {vt,t,u,vt}. Hence, the value of the evaluations is represented as a tuple of four rational numbers, each one of them between 0 and 1 and summing exactly one. Fig. 1 shows the possible values of this representation. An example of a reputation theory for the agent *i* could be:

$$DE_i(b_1, seller, [1, 0, 0, 0], t_1)$$
$$DE_i(b_1, seller, [0, 1, 0, 0], t_2)$$
$$DE_i(b_2, seller, [0, 0, 0, 1], t_3)$$
$$DE_i(b_2, seller, [0, 0, 0, 1], t_4)$$
$$Comm_i(u_1, DE_{u_1}(b_1, seller, [1, 0, 0, 0], t_x), t_5)$$
$$Comm_i(u_2, DE_{u_1}(b_1, seller, [0, 1, 0, 0], t_y), t_6)$$

Agent *i* is able to infer image predicates from the theory above. The *trust* measure that the model provides, in terms of L_{rep}, coincide with the concept of image, because agents accept the measure as *true*. Then, using the transformation shown in Fig. 1, the following image predicates can be inferred:

$$Img_i(b_1, seller, [0.5, 0.5, 0, 0])$$
$$Img_i(b_2, seller, [0, 0, 0, 1])$$

3. Communicated social evaluations and their reliability

3.1. Preliminaries

The problem regarding communicated social evaluations that the subjective notion of reputation brings, is the same as for any rhetorical construct that depends on internal elements that are private. Let us consider a very simple example:

i: *How is John as a car driver?*

j: *He is a very good driver*

i: *Why?*

j: *Well, Emma told me that, and she is a good informer*

i: *Oh! for me, Emma is very bad as informer!*

In the previous example, should agent *i* consider the information sent by *j* saying that *John* is a *very good* driver? Notice that *j* is justifying her opinion with a previous communication from *Emma*, which she thinks is a good informer. But it contradicts an information that *i* considers valid. For *i*, the information is not reliable, even when *j* may be totally honest.

When talking about social evaluations and reputation models, usually the model already handles possible inconsistent knowledge. Different opinions referring to the same target agent may be totally contradictory, and the agent integrates and aggregates the information in order to achieve a consistent mental state. Determining whether a piece of information is acceptable in a possibly inconsistent knowledge base has been faced in argumentation theory. In this field, each piece of information is justified by the elementary elements from which it has been inferred, the so called arguments. Then, two arguments can *attack* each other, indicating that the information supporting them would be inconsistent if they are both accepted at the same time.

3.2. Characterizing the problems behind reliability measures

Having introduced the reputation language, we can illustrate more precisely the kind of problems we deal with in this paper, and the characteristics of the proposed system. We start with a very simple example. Let *i*, *j* be two agents with their respective reputation theories and reputation models $\langle \Delta_i, \vdash_i \rangle$ and $\langle \Delta_j, \vdash_j \rangle$. Let us consider that agent *i* has a *VG* image of *John* as a *car_seller* (with a maximum reliability), so $\Delta_i \vdash_i Img_i(John, car\_seller, VG)$. When *i* communicates such information to *j* at time *t*, *j* updates its reputation theory with a new communication:

$$\Delta_j \cup \{Comm_j(i, \ulcorner Img(John, car\_seller, VG) \urcorner, t)\}$$

Let us assume that *i* inferred the image of *John* as a *car_seller* from (1) a communication from *Alice* and (2) the very good reputation (according to *i*) of *Alice* as *informer*:

$$(1)\ Comm_i(Alice, \ulcorner Img_{Alice}(John, car\_seller, VG) \urcorner, t)$$

$$(2)\ Rep_i(Alice, informer, VG)$$

Also, assume that *j* has a very different opinion about the reputation of *Alice* as *infomer*, a very bad reputation indeed. Specifically, $\Delta_j \vdash_j Rep_j(Alice, informer, VB)$. With this scenario, at least one question arises. Should *j* update its reputation theory with the original communication from *i*?

We argue that the communicated information from *i* is not reliable for *j* in this example. Without the analysis of the internal elements, such a situation is impossible to detect, and the effects of including *i*'s communication in *j*'s reputation theory can be devastating for *j*. Agents use social evaluations to decide what to do. It may happen that *i*'s communication helps *j* choose *John* as a *car_seller* when *j* wants to buy a car. If the direct interaction with *John* does not go well, several things may occur:

1. Direct experiences are costly. Probably *j* has bought a car before noticing that it was not good.
2. *j* may generate a bad image of *i* as informer, which can lead to *j* not considering any future communications from *i*, even when *i*, according to *j*'s knowledge, was honest.
3. Also *j* may spread bad reputation of *i* as informer, and thus collide with the opinion of other members of the society that are aligned with *i*. Consequently such members may retaliate against *j* [4].

All the previous situations could be avoided if *j* has the capability to decide whether the piece of information is reliable enough, not based on the reliability measure that *i* assigns, but on the internal elements that *i* uses to justify the communicated social evaluation and that *j* can check. Furthermore, our approach makes it more difficult to intentionally lie, since a potential liar should know beforehand what the recipient knows, and build the argument accordingly to it. In current approaches, a liar agent can put a very high reliability value in the communicated social evaluation to introduce noise in the recipient agent.

To allow agents to analyze the justifications, we propose a protocol that implements a dialectical process between the agents. Intuitively, both the source and the recipient agents, following a well-defined protocol, can exchange at each turn a justified social evaluation (argument) that counterargues (attacks) some of the arguments uttered by the other agent. At the end of the process, the recipient agent holds a tree of arguments that can be used to decide whether the original communication from the source agent is reliable, and update its reputation theory accordingly. The technical details to design such a protocol and the posterior analysis are taken from the field of computation argumentation, which has proposed frameworks and methods to deal with similar situations. We have taken some of these concepts and tools and adapted them to confront the peculiarities that reputation information and our scenarios have. We highlight just two:

The attacks are graded: In the previous example, *j* holds a very different opinion of the reputation of *Alice* as *informer* than *i* has, *very bad* (VB) against *very good* (VG) respectively. However, this would note the case if *j* thinks that the reputation

of *alice* as *informer* is *good* (G), so $\Delta_j \vdash_j Rep(Alice, informer, G)$. The *attack* should be considered weaker in the latter case. Our framework handles graded attacks by assuming that each agent has a distance function $\ominus : G \times G \rightarrow \mathbb{Q}$ over the totally order set $M = \langle G, \leq \rangle$ which is used to represent the values of the social evaluations (see Section 2).

Heterogeneity of the agents: Even when agents use the same language to talk and reason about reputation, they may use different reputation models. Usually, an argument is defined as a pair composed of a conclusion and a set of elements that have been used to infer such conclusion (supporting set). The conclusion is the element that is being justified by the supporting set. If agents use different inference rules, the supporting set must include enough information to reconstruct the reasoning path followed by the agent that has built the argument. This could also be *easily* done by sending the exact inference rules of the reputation model in the arguments, but it would violate the privacy of the agents and therefore is not an option. Instead, our framework provides an intermediate solution. We define a very simple inference consequence relation $\vdash_{arg}$ that all agents must know, and specify a transformation that agent should use to build arguments using $\vdash_{arg}$. From $\langle \Delta_i, \vdash_i \rangle$ and $\langle \Delta_j, \vdash_j \rangle$, we move to $\langle \Gamma_i, \vdash_{arg} \rangle$ and $\langle \Gamma_j, \vdash_{arg} \rangle$, where Γ_i and Γ_j are argumentative theories built from their respective reputation theories and reputation models. Argumentative theories contain all the elements from their respective reputation theories, and simple implication rules that simulate inference steps performed by their respective reputation model, without indicating how they were performed internally.

The protocol allows agents to construct trees of arguments with their respective attacks. We provide an acceptability semantics, a mechanism for deciding whether the information from the source agent can be considered reliable for the recipient. We can do that because the argumentation framework we instantiate introduces the concept of *inconsistency budgets* [10]. Intuitively, inconsistency budgets indicate the *amount* of inconsistency that an agent can (wants to) tolerate. For instance, in the previous example where $\Delta_j \vdash_j Rep(Alice, informer, G)$, agent *j* may consider that the difference between *G* and *VG* is small enough to accept that they are not contradictory, even when that might not be the case for another agent. Agents *autonomously* decide the strength of a given attack according to their own distance function and therefore to which extent they can accept inconsistencies.

The next section formally describes: (1) how agents build arguments; (2) how agents construct an argumentative theory from a reputation theory; (3) how such arguments influence each other and with which strength; and (4) how the recipient agent can decide whether a piece of communicated information is reliable.

4. The reputation argumentation framework

Our approach suggests that agents use argumentation techniques to decide whether a piece of information can be considered reliable or not. For this, we need to define an argumentation framework for reputation-related concepts. First, we specify the notion of argument, the construct of arguments, and how they influence each other. Second, we define L_{arg}, a language based on L_{Rep} to write argument sentences, and the consequence relation $\vdash_{arg}$ associated with the language and used to build arguments. We also give an acceptability semantics, indicating under which conditions, an agent would *accept* a given communicated social evaluation as reliable.

Definition (Argument)**.** A formula $(\Phi{:}\alpha) \in wff(L_{arg})$ when $\alpha \in wff(L_{Rep})$ and $\Phi \subseteq wff(L_{Rep})$. Intuitively, we say that the set Φ is the supporting set of the argument, and α its conclusion. It indicates that α has been deduced from the elements in Φ.

The validity of a given well-formed argument must be contextualized in an argumentation theory, a set of elementary argumentative formulas, called *basic declarative units* (BDU). We adapt the following definition from [11]:

Definition (Argumentative theory)**.** A *basic declarative unit* (BDU) is a formula $(\{\alpha\}{:}\alpha) \in wff(L_{arg})$. Then, a finite set $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ is an argumentative theory iff each γ_i is a BDU.

From an argumentative theory Γ, we can now define how arguments are constructed. For this we use the inference relation $\vdash_{arg}$, characterized by the deduction rules *Intro-BDU*, *Intro-AND* and *Elim-IMP* (Fig. 2). Rule *Intro-BDU* allows the introduction of a basic declarative unit from the argumentative theory. It is necessary to ensure completeness (proposition 4.1), permitting each BDU formula from the theory to be deduced via $\vdash_{arg}$. Rule *Intro-AND* permits the introduction of conjunctions. Finally, rule *Elim-IMP* performs the traditional *modus ponens*.

Definition (Valid argument and subargument)**.** Let $(\Phi{:}\alpha) \in wff(L_{arg})$ and let Γ be an argumentative theory. We say that $(\Phi{:}\alpha)$ is a valid argument in the bases of Γ iff $\Gamma \vdash_{arg} (\Phi{:}\alpha)$. Also, we say that a valid argument $(\Phi_2{:}\alpha_2)$ is a subargument of $(\Phi{:}\alpha)$ iff $\Phi_2 \subset \Phi$.

As mentioned earlier, each agent *i* constructs its argumentative theory Γ_i in order to build arguments. This argumentative theory is based on the reputation information that *i* has, characterized with the tuple $\langle \Delta_i, \vdash_i \rangle$. Assuming that $\vdash_i$ is defined by a finite set of natural deduction rules $\{\vdash_{i_1}, \ldots, \vdash_{i_m}\}$,

$$\text{Intro-BDU:}\ \frac{}{(\{\alpha\}:\alpha)} \qquad \text{Intro-AND:}\ \frac{(\Phi_1:\alpha_1),\ldots\ (\Phi_n:\alpha_n)}{(\bigcup_{i=1}^{n}\Phi_i:\alpha_1,\ldots,\alpha_n)}$$

$$\text{Elim-IMP:}\ \frac{(\Phi_1:\alpha_1,\ldots,\alpha_n\rightarrow\beta)\qquad(\Phi_2:\alpha_1,\ldots,\alpha_n)}{(\Phi_1\cup\Phi_2:\beta)}$$

Fig. 2. Deductive rules for the consequence relation $\vdash_{arg}$.

- For all $\alpha\in\Delta_i$ then $(\{\alpha\}:\alpha)\in\Gamma_i$. That is, all ground elements from the reputation theory are BDU in the argumentative theory.
- For all $\alpha_1,\ldots,\alpha_n$ s.t. $\Delta_i\vdash_i\alpha_k$ where $1\leq k\leq n$, if there exists m s.t. $\alpha_1,\ldots,\alpha_n\vdash_{i_m}\beta$, then $(\{\alpha_1,\ldots,\alpha_n\rightarrow\beta\}:\alpha_1,\ldots,\alpha_n\rightarrow\beta)\in\Gamma_i$. This construct introduces every instantiated deductive step as a rule in the form of a basic declarative unit. For instance, if $\alpha,\beta\vdash_{i_2}\gamma$, the argumentative theory will include the BDU formula $(\{\alpha,\beta\rightarrow\gamma\}:\alpha,\beta\rightarrow\gamma)$.

The following proposition is easy to prove and establishes the completeness of the deductive system.

Proposition 4.1. *Let $\langle\Delta_i,\vdash_i\rangle$ be the reputation information associated with agent i, and Γ_i its argumentative theory. If $\Delta_i\vdash_i\alpha$, then there exists an argument $(\Phi:\alpha)$ such that $\Gamma_i\vdash_{arg}(\Phi:\alpha)$.*

4.1. Argument interactions

We have explained how agents construct their argumentative theory from their reputation information, and how from such theory they can build arguments using $\vdash_{arg}$. In this subsection we detail how arguments generated from different agents influence one other. Unlike argumentation systems that are used as theoretical reasoning processes this does not imply necessary that the attack relation must be symmetric. Differently from argumentation systems used as theoretical reasoning processes to analyze the possible inconsistencies that a single agent may hold, our framework is designed to be part of a dialectical process, where attacks are produced only from arguments sent by other agents. Note that in our framework attacks may or may not be symmetric.

To specify the *attack* relationship among arguments, we define first the binary relation $\cong$ between L_{rep} predicates. Let α,β be well-formed non-ground formulas from L_{Rep}. Then, $\alpha\cong\beta$ iff $type(\alpha)=type(\beta)$, $\alpha.target=\beta.target$, $\alpha.context=\beta.context$ and $\alpha.value\neq\beta.value$. We can see that $\cong$ is symmetric but not reflexive nor transitive. For instance, $Rep(i,seller,VB)\cong Rep(i,seller,G)$, but $Rep(i,seller,VB)\ncong Img(i,seller,G)$ and $Rep(i,seller,VB)\ncong Rep(i,buyer,VG)$.

Definition (Attack between arguments)**.** Let $(\Phi_1:\alpha_1)$, $(\Phi_2:\alpha_2)$ be valid arguments in the bases of Γ. We say that $(\Phi_1:\alpha_1)$ *attacks* $(\Phi_2:\alpha_2)$ iff $\exists(\Phi_3:\alpha_3)$ subargument of $(\Phi_2:\alpha_2)$ s.t. $(\alpha_1\cong\alpha_3)$.

We want also to quantify the strength of the attack. Let $a=(\Phi_1:\alpha_1)$ be an argument that attacks $b=(\Phi_2:\alpha_2)$. Then, by definition, a $(\Phi_3:\alpha_3)$ subargument of $(\Phi_2:\alpha_2)$ s.t. $(\alpha_1\cong\alpha_3)$ exists. The strength of the attack is calculated through the function w as $w(a,b)=\alpha_1.value\ominus\alpha_3.value$, where $\ominus$ is a binary function defined over the domain of the representation values used to quantify the evaluations (the total ordered set $M=\langle G,\leq\rangle$). For instance, if $M=\langle[0,1]\cap\mathbb{Q},\leq\rangle$, we can define $\ominus(x,y)=|x-y|$. In this case, 1 is the strongest attack. If $M=\langle\{VB,B,N,G,VG\},\leq_s\rangle$, we could first assign each label a number: $f(VB)=0, f(B)=1$, ..., and then, $\ominus(x,y)=|f(x)-f(y)|$. In this case, the strongest attack is quantified with 4. $\ominus$ implements a *difference* function among the possible values.

The previous attack definition does not consider attacks between direct experiences nor communications. This means that discrepancies at this level cannot be argued, even when they are completely contradictory. Yet, this is justified by the fact that, in our framework, ground elements are not generated from any other piece of information. Thus, a communicated ground element should be introduced directly into the reputation theory. Obviously, the language could be extended to capture the elementary predicates that compose direct experiences (contracts, fulfillments etc.). Again though, we think that sharing this low level information would violate the privacy of the agents.

4.2. Deciding about the reliability

At this point, agents can build arguments, determine when their arguments attack arguments from other agents (and vise versa), and assign a strength to these attacks. However, we are still missing how agents can decide when to accept a given argument, considering that they will have a weighted tree of arguments where each node is an argument and each edge represents the strength of the attack. For this, we instantiate a weighted version of the classic Dung abstract argumentation framework [12], and use an acceptability semantics defined for this framework.

Dung's framework is defined as follows:

Definition (Abstract argumentation framework)**.** An abstract argument system (or argumentation framework) is a tuple $AF = \langle A, R\rangle$ where A is a set of arguments and $R \subseteq A \times A$ is an attack relation. Given $a, b \in A$, if $(a, b) \in R$ (or aRb), we say that a attacks b. Let $S \subseteq A$, and $a, b \in A$ then

- S is *conflict-free* iff $\nexists a, b \in S$ s.t. aRb.
- An argument b is *acceptable* w.r.t. the set S iff $\forall a \in A$, if aRb then $\exists c \in S$ s.t. cRa.
- S is *admissible* if it is conflict-free, and each argument in S is acceptable w.r.t. the set S. Also, S is a preferred extension iff it is maximal w.r.t. the set inclusion.
- An argument b is *credulously accepted* iff it belongs to at least one preferred extension.

This abstract framework does not consider strength in the attacks. Work from Dunne et al. [10] extends Dung's framework with weights.

Definition (Weighted abstract argumentation framework)**.** A weighted argument system is a triple $AF_w = \langle A, R, w\rangle$ where $\langle A, R\rangle$ corresponds to a Dung's argumentation framework, and $w : R \to \boldsymbol{R}_{>}$ is a function that assigns weights to each attack relation.[4]

The semantics of w gives a pre-order between possible inconsistencies. Let $a_1, b_1, a_2, b_2 \in A$ where a_1Rb_1 and a_2Rb_2, if $w((a_1, b_1)) < w((a_2, b_2))$ means that accepting both a_1 and b_1 is *more consistent* than accepting both a_2 and b_2. This leads to the definition of inconsistency budgets and $\beta-$solutions (β s.t. $\beta \in \boldsymbol{R}_{\geq}$). Intuitively, a β-solution is a solution of the unweighted Dung's framework in which the *amount* of inconsistency (calculated through the sum of the weights of the attacks) is lower or equal to β. Formally:

Definition (β-solutions [10])**.** Given $AF_w = \langle A, R, w\rangle$, a solution $S \subseteq A$ is a $\beta-$solution if $\exists T \in sub(R, w, \beta)$ s.t. S is a solution of the unweighed system $AF = \langle A, R\backslash T\rangle$. Function *sub* returns a set of subsets of R in which the weights sum up to a maximum of β: $sub(R, w, \beta) = \{T | T \subseteq R \text{ and } (\sum_{r\in T} w(r)) \leq \beta\}$

We use a credulous semantics for the acceptance of reliable information, although alternatively, other semantics could be used. Credulous semantics ensures that at least, the argument belongs to one preferred extension, which is what we are looking for. In the weighted version, we can define that, given $AF_w = \langle A, R, w\rangle$, an argument $b \in A$ is credulously accepted if it belongs to at least one β-preferred extension, so, if $\exists T \in sub(R, w, \beta)$ s.t. $b \in S$, and S is a preferred extension of the Dung's framework $AF = \langle A, R\backslash T\rangle$.

We can instantiate now the weighted argument system by using the constructs defined in this section. Let Γ be an argumentative theory as defined in this section. We define:

- $C(\Gamma) = \{(\Phi : \alpha) | \Gamma \vdash_{arg} (\Phi : \alpha)\}$, the set of all valid arguments that can be deduced from Γ.
- $R(\Gamma) = \{((\Phi_1 : \alpha_1), (\Phi_2 : \alpha_2)) | (\Phi_1 : \alpha_1)$ attacks $(\Phi_2 : \alpha_2)$ and $(\Phi_1 : \alpha_1) \in C(\Gamma)$ and $(\Phi_2 : \alpha_2) \in C(\Gamma)\}$, the set of all possible attack relations between the arguments in $C(\Gamma)$.

Then, we can describe the instantiation:

Definition (Reputation argument framework)**.** The reputation argument system for the argumentative theory Γ is defined as $AF_\Gamma = \langle C(\Gamma), R(\Gamma), w\rangle$, where $w : R(\Gamma) \to \boldsymbol{R}$ is the strength function as defined above using the $\ominus$ difference function.

This finishes the definition of the reputation argument system. The idea is that each agent will be equipped with its own argumentation reputation system, and will incrementally add the arguments issued by the other agent. Intuitively, if the argument that justifies the original communicated social evaluation belongs to a preferred extension of the recipient agent, the latter will introduce the social evaluation into its reputation theory.

5. The dialog protocol

A dialog between two parties that can be seen as a game in which each agent has an objective and a set of legal movements (illocutions) to perform at each turn. Walton *et al.* in [13] state several types of dialogs depending on the participants' goals. In our case, we model a special kind of *information-seeking* dialog. The goal of the game then is to see whether the opponent (OPP) can *accept* reasonably the inquiring information from the proponent (PRO).

[4] Following the notation in [10], we write $\boldsymbol{R}_{>}$ and $\boldsymbol{R}_{\geq}$ to refer to the set of real numbers greater than 0 and greater or equal to 0 respectively.

Table 1
Possible moves of the dialog game at turn k. The function $supp(b)$ returns the supporting set of b.

$\text{counter}^k_{PRO}(b)$	**Precondition** (1) k is even, $b \in C(\Gamma_{PRO} \cup X^{k-1}_{OPP})$ and b has not been issued yet (2) $\exists r \in \mathbb{N}$ s.t. $1 \leq r < \lvert S^{k-1}\rvert$, r is odd and $(b, S^{k-1}_r) \in R(\Gamma_{PRO} \cup X^{k-1}_{OPP})$ (3) $\nexists \gamma \in C(\Gamma_{PRO} \cup X^{k-1}_{OPP})$ s.t. $(\gamma, S^{k-1}_t) \in R(\Gamma_{PRO} \cup X^{k-1}_{OPP})$ where $r+1 \leq t < \lvert S^{k-1}\rvert$ and t is odd **Postcondition** (i) $X^k_{PRO} = X^{k-1}_{PRO} \cup BDU(supp(b))$ (ii) $X^k_{OPP} = X^{k-1}_{OPP}$ (iii) $S^k = \langle S^{k-1}_0, \ldots, S^{k-1}_r, b\rangle$
$\text{counter}^k_{OPP}(b)$	**Precondition** (1) k is odd, $b \in C(\Gamma_{OPP} \cup X^{k-1}_{PRO})$ and b has not been issued yet (2) $\exists r \in \mathbb{N}$ s.t. $0 \leq r < \lvert S^{k-1}\rvert$, r is even and $(b, S^{k-1}_r) \in R(\Gamma_{OPP} \cup X^{k-1}_{PRO})$ (3) $\nexists \gamma \in C(\Gamma_{OPP} \cup X^{k-1}_{PRO})$ s.t. $(\gamma, S^{k-1}_t) \in R(\Gamma_{OPP} \cup X^{k-1}_{PRO})$ where $r+1 \leq t < \lvert S^{k-1}\rvert$ and t is even **Postcondition** (i) $X^k_{PRO} = X^{k-1}_{PRO}$ (ii) $X^k_{OPP} = X^{k-1}_{OPP} \cup BDU(supp(b))$ (iii) $S^k = \langle S^{k-1}_0, \ldots, S^{k-1}_r, b\rangle$ (or $\langle S^{k-1}_0, b\rangle$ if $r = 0$)

We use the argumentation framework defined in the previous sections to give semantics to the dialogs. The key is that each agent participating in the dialog will use its own argument framework to deal with possible inconsistencies. It is important to notice that agents do not have access to the set of arguments of the other agents. They incorporate such knowledge from the exchange of illocutions uttered in the dialog.

Let PRO and OPP be the proponent and the opponent agents engaged in the dialog respectively. Following a similar approach used in [14], both agents are equipped with a reputation argument system:

$$AF_{PRO} = \langle C(\Gamma_{PRO} \cup X_{OPP}), R(\Gamma_{PRO} \cup X_{OPP}), w_{PRO}\rangle$$
$$AF_{OPP} = \langle C(\Gamma_{OPP} \cup X_{PRO}), R(\Gamma_{OPP} \cup X_{PRO}), w_{OPP}\rangle$$

where Γ_{PRO}, Γ_{OPP} are the argumentative theories of agents PRO and OPP, which are private. w_{PRO} and w_{OPP} are the weight functions of agents PRO and OPP. Finally, X_{PRO} (X_{OPP}) is the set of BDU from the arguments that results from the proponent's (opponent's) issued arguments. Both X_{PRO} and X_{OPP} are public and are the result of the exchange of arguments. This allows the agents to recognize and reconstruct arguments from the other agent. As for the state of our dialog protocol, we give a definition inspired by [15]:

Definition (State of the dialog)**.** A state of a dialog at the kth turn (where $k \geq 0$) is characterized by the tuple $\langle S^k, X^k_{PRO}, X^k_{OPP}\rangle^k$ where $S^k = \langle S^k_0, \ldots, S^k_t\rangle$ is the ordered set of arguments that represents a single dispute line. A dispute line is a finite sequence of arguments $a_0, \ldots, a_n$ where $\forall l$ s.t. $1 \leq l < n$, a_{l+1} attacks a_l. X^k_{PRO}, X^k_{OPP} are the public sets of BDU formulas of the proponent and the opponent respectively at turn k, incrementally built after each argument exchange and that are public.

The proponent is the initiator of the dialog and issues the argument $a = (\Phi{:}\alpha)$. The initial state at turn 0 is then characterized by $\langle\langle a\rangle, BDU(\Phi), \{\}\rangle^0$. The function $BDU(X)$ returns the set of elements from X as a BDU formula. So, if $\alpha \in X$, then $\{\alpha\}{:}\alpha \in BDU(X)$. The possible types of movements are summarized in Table 1, where we include preconditions and postconditions:

The proponent can perform the movement $counter^k_{PRO}(b)$ when the turn k is even (1). Of course, b should be a valid argument built from its argumentative theory and the BDU from the previous exchange of arguments ($C(\Gamma_{PRO} \cup X^{k-1}_{OPP})$) (1). We also require that b attacks some of the arguments of the current dispute line that the opponent has issued (so, in an odd position) (2). When this occurs, we also want to ensure that the proponent cannot attack any other argument issued by the opponent later than the one being attacked (3). Once the illocution is submitted, the effects in the dialog state are also described in Table 1. First, the set X_{PRO} is updated with the supporting set of the argument b (i). Notice that in the way we define the construction of arguments (see Section 4) the supporting set only contains BDU. Thus, since this set is added to the argumentative theory of the opponent, it is able also to recognize the argument and attack it if necessary. Moreover, when an argument of the dispute line is attacked at point r of the dispute line, the dialog starts a new dispute line from that point (iii).

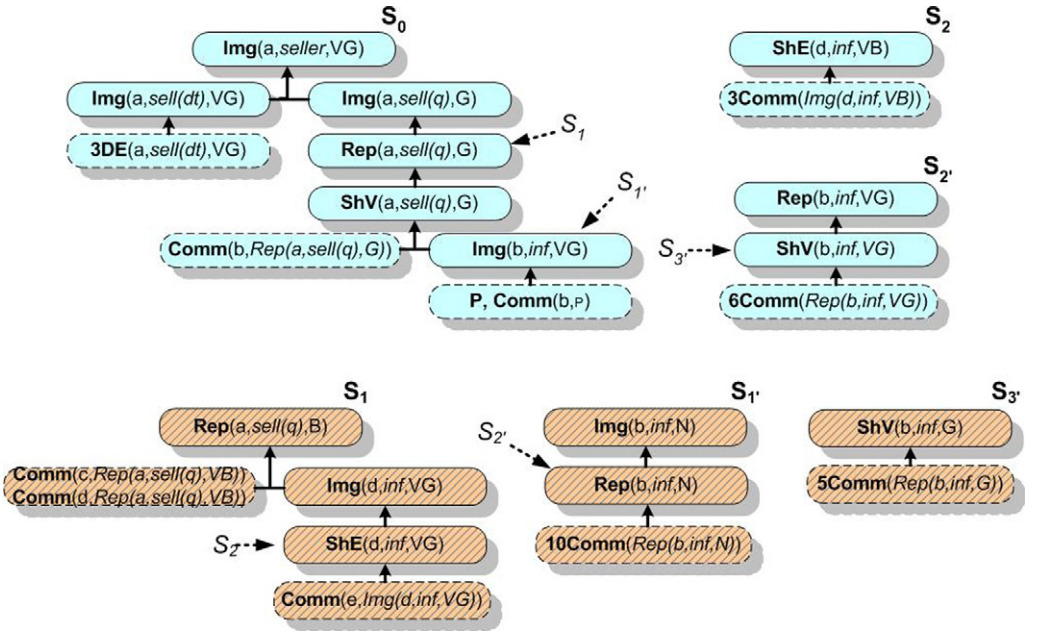

Fig. 3. Arguments uttered by the proponent (S_0,S_2,$S_{2'}$) and the opponent (S_1,$S_{1'}$,$S_{3'}$) in the example respectively. Dotted arrows indicate attack. For the sake of clarity, we omit the notation $\lceil\cdot\rceil$. $nComm(\cdot)$ and $nDE(\cdot)$ indicates that the agent holds n communications and n direct experiences respectively.

The opponent can submit counterarguments by sending the illocution $counter^k_{OPP}(b)$ with symmetric effects as explained in the previous paragraph. In this case, k must be odd. A dialog finishes when there are no possible movements.

The winner is the last participant who makes a move. Hence, if the number of moves is even, the winner is the proponent. If the number of moves is odd, the opponent wins. This protocol is a simplification of a TPI-Dispute (*two-party immediate response dispute*) and instantiates a protocol described in [14]. From there, the following proposition can be deduced:

Proposition 5.1. *Let AF_{PRO} and AF_{OPP} be the argument frameworks of the participants of a dialog. When the game is finished and the proponent is the winner, the original argument $a = (\Phi{:}\alpha)$ belongs to a* 0*-preferred extensions of AF_{OPP}.*

This means that the argument a is credulously accepted by the opponent. Therefore, the conclusion α can be introduced into the reputation theory of the opponent. It is easy to prove that the dialog incrementally builds a dialectical tree, which has been proved to construct admissible extensions [14]. This is one of the advantages of using argumentation frameworks. Agents do not need to generate all possible arguments to determined the state of a given argument. It is enough to generate a dialectical tree, which is what in fact, the protocol we have defined perform.

If OPP wins, OPP cannot find a 0-preferred extension that includes the argument a. In this case, OPP could choose not to update its reputation theory. However, depending on its tolerance to inconsistencies, OPP can find a 1-preferred extension that includes argument a, or even a 2-preferred extension. By increasing the inconsistency budget, the original argument may become acceptable, and thus the communicated social evaluation considered reliable. This might be seen equivalent to the threshold that some reputation models that manage reliability measures use to accept communicated information. The difference is that contrary to the measure being calculated by the source agent, in our approach, the reliability is computed by the recipient, who assigns strengths that can be different from the source. Algorithm 1 formalizes the procedure we have just described. In the next subsection we provide an example that shows the use of the protocol and the inconsistency budgets.

5.1. An example

We want to finish this section by showing a simple example. Here, agent i (the proponent) sends the first communication to j (the opponent). The arguments they build are shown in Fig. 3. In the domain, we have the context *seller*, composed of *sell(q)* (quality dimension of the sold products) and *sell(dt)* (delivery time of the product). We use the L_{rep} language taking M as $\langle\{VB, B, N, G, VG\}, \leq_s\rangle$. Also, the context *Inf* is used and stands for *informant*. For instance, the argument S_0 (Fig. 3) indicates that the agent has a VG (very good) image of a as a seller, because of the images it has about a taking into account the quality of the products (sell(q)) and the delivery time (sell(dt)) are G and VG respectively. The latter is justified because it had three direct experiences with a resulting in a very good delivery time (VG), and so on. In the figure, elements in dot lines belong to the ground elements of the argumentation theory of i. Arrows represent implication relation which are also in the argumentative theory. For instance, there is an implication relation in the theory that says:

Algorithm 1: Reputation Theory Update (for agent j)

Data: Agent i, j
Data: Argument $\Phi{:}\alpha$ (sent by i)
Data: Reputation Information $\langle \Delta, \vdash_R \rangle$
Data: Inconsistency Budget b
Result: Δ_{res} (Reputation Theory Updated)
$\Gamma_j \leftarrow$ Argumentative Theory from $\langle \Delta, \vdash_R \rangle$;
$AF_j \leftarrow \langle C(\Gamma_j), R(\Gamma_j), w_j \rangle$ /*The argument framework of j*/;
$\langle winner, X_i \rangle \leftarrow$ dialogGame(AF_j,i, $\langle \langle \alpha \rangle, \Phi, \{\} \rangle^0$); **if** $winner = i$ **then**
  $\Delta_{res} \leftarrow \Delta \cup \{Comm(i, \alpha)\}$ /*i wins, then j updates its reputation theory*/;
else
  if $\Phi{:}\alpha$ *is acceptable w.r.t.* $\langle C(\Gamma_j \cup X_i), R_j, w_j \rangle$ *and budget* b **then**
    $\Delta_{res} \leftarrow \Delta \cup \{Comm(i, \alpha)\}$ /*With inconsistency budget b, j accepts also the argument*/;
  else
    $\Delta_{res} \leftarrow \Delta$ /*Agent j rejects the argument*/;
  end
end

$$Img(a, sell(dt), VG), Img(a, sell(q), G) \rightarrow Img(a, seller, VG)$$

and another saying that

$$Rep(a, sell(q), G) \rightarrow Img(a, sell(q), G)$$

In Fig. 3 we show arguments and sub-arguments already instantiated to facilitate the reading.

The next table shows the illocutions that the agents exchange. The column *Dispute Line* shows the state of the dispute line.

Action	Dispute Line
–	$S^0 = \{S_0\}$
$\text{counter}^1_{OPP}(S_1)$	$S^1 = \{S_0, S_1\}$
$\text{counter}^2_{PRO}(S_2)$	$S^2 = \{S_0, S_1, S_2\}$
$\text{counter}^3_{OPP}(S_{1'})$	$S^3 = \{S_0, S_{1'}\}$
$\text{counter}^4_{PRO}(S_{2'})$	$S^4 = \{S_0, S_{1'}, S_{2'}\}$
$\text{counter}^5_{OPP}(S_{3'})$	$S^5 = \{S_0, S_{1'}, S_{2'}, S_{3'}\}$

In the first move, the opponent (OPP) utters the argument S_1 which attacks the original S_0. S_1 has the conclusion formula $Rep(a, sell(q), B)$ and attacks the subargument of S_0 that has as a conclusion $Rep(a, sell(q), G)$. The strength is calculated applying the function $\ominus$ on the values of the predicates. In this case, $\ominus(B, G) = 2$. In the next move, the proponent (PRO) attacks S_1 by sending S_2 (strength = 4). At this point, we assume that OPP cannot attack S_2, but it can attack again the original S_0. In movement 3, OPP sends the argument $S_{1'}$ to attack S_0 (strength = 2). Notice that the dispute line has changed. Then, the proponent counterargues $S_{1'}$ by sending $S_{2'}$ (strength = 2). Finally, OPP finishes the game at movement 5 by issuing $S_{3'}$, which attacks $S_{2'}$ (strength = 1).

The opponent wins the game. This means that OPP considers the initial information from PRO unreliable . The dialectical tree after the game is shown in Fig. 4. With this tree, OPP cannot construct an admissible set that includes S_0, and thus cannot accept it. But this is only true when OPP takes an inconsistency budget of 0. As soon as it tolerates a budget of 1, the result changes. Now, the set $\{S_0, S_2, S_{2'}, S_{3'}\}$ is a 1-preferred extension and S_0 becomes acceptable. At this point, OPP could update its reputation theory, considering that the information is reliable enough.

It is important to recall again that this does not mean that j accepts the conclusion of S_0 in the classical logical sense. It means that j adds it in the reputation theory as a communication. The reputation model will be in charge of updating the corresponding social evaluation taking into account other information like for instance, the reliability of i as informant.

6. Experimental results

In the previous sections we have presented the theoretical development of the reputation argumentation framework and have discussed the reasons why we need such a system, and under which theoretical conditions the framework can be used to argue about reputation concepts. We have shown the completeness of the deductive system (proposition 4.1) and the

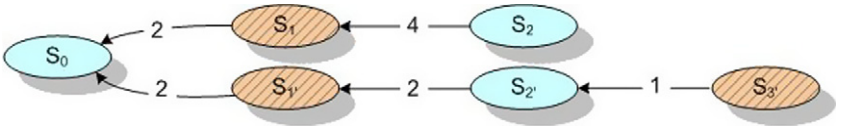

Fig. 4. The dialectical tree of agent *j* after the game. Arrows represent attack relation and the labels indicate the strength of the attacks.

correctness of the argumentation-based protocol (proposition 5.1). Nevertheless, under the assumption of self-interested agents that try to maximize their utility functions, it is not always clear whether it is worth engaging in a dialog protocol.

Firstly, there is an obvious trade off between time and cost, which in our framework is related strictly to *the number of exchanged messages* and the *achieved accuracy* respectively. In this sense,

1. When the cost of a bad interaction is higher than the cost of messaging (or waiting time), the accuracy becomes a crucial issue, and argumentation can help.
2. On the contrary, when the cost of messaging dominates the potential failure of an interaction, for sure argumentation is not a good solution. For example, application involving wireless sensor networks where messaging is a critical issue due to the high energy cost, argumentation-based protocols are not the best solutions.

We focus on the former, and assume that the cost of messaging is not relevant in comparison to the cost of a bad interaction.[5]

Also, notice that the reputation argumentation framework is a complement attached to an existing reputation model, which in general, is already pretty accurate. In fact, there is no guarantee that the use of our argumentation protocol significantly improves the accuracy of the agents. In this section we present the results of a set of simulations that explore some of the relevant parameters that we consider crucial. We empirically validate that the argumentation-based protocol for social evaluation exchange significantly improves the accuracy of the agents when modeling the behavior of others.

The simulations should be considered a proof-of-concept environment that proves that in the scenario described below the use of argumentation is useful. Even when we would require a more complete set of experiments to completely validate the utility of the argumentation-based protocol, interesting conclusions can be extracted, and of course can be extrapolated to other scenarios.[6]

6.1. Description of the simulations

Similar to [17–19], we consider an scenario with buyers and sellers. Sellers offer products with constant quality (q) and deliver them with a constant delivery time (dt). Buyers are endowed with a reputation model and evaluate sellers in the role *seller*, which is composed of *sell(q)* (quality dimension of the sold products) and *sell(dt)* (delivery time). We also consider the context *Inf*, which stands for *informant*.

We consider that different buyers can evaluate sellers in different ways (they have different goals). To simplify, some buyers only take into account good sellers according to the quality of the products (QBuyers), while others, according to the delivery time (DTBuyers). This resembles the different goals of the agent. The key point is that initially, buyers communicate social evaluations only regarding the role *seller*. So, a *good seller* for agent *A* is not necessary good for agent *B*.

Under standard settings, the introduction of information from unaligned buyers may bias the accuracy of the reputation models, while when using argumentation, such information can be filtered the accuracy improved. This evidence clashes with the idea that to argue, both the source and the recipient agents must have some knowledge about the environment, but not *too much*. If agents do not have any information (or few), no argumentation is possible. On the opposite, if agents have already a lot of information that includes a high number of direct trades, agents will not be able to respond, since direct experiences cannot be attacked in our framework. To parametrize the former situation we include a bootstrap phase, where agents explore the environment without arguing. We do not let agents trade directly with all the agents, only with a subset of them. In concrete, our simulations have the following phases:

- *Bootstrap phase:* It is used to endow the buyers with some knowledge about the environment (that is, other sellers and buyers). At each turn, each buyer performs two tasks: (1) it chooses a seller randomly, buying from it, and (2) it sends a communication of an image predicate regarding a random seller or buyer (in the latter in the role of *Inf*) to a random buyer agent. No argumentation is present in this phase. The number of direct trades and messages sent at each turn can be parameterized. In our simulations we allow each agent one direct trade and one message per turn.

  To parametrize the fact that agents do not have too much information in terms of direct trades, a percentage of the buyers (*pctBuyers-Bootstrap*) can only perform direct trades with a percentage of sellers (*pctSellers-Bootstrap*). The other

[5] Even when the cost of messaging is null, if the cost of interacting is very low there is no motivation for the agents to exchange information. Experimental evidence of this can be found in [16].

[6] The simulations were performed in JAVA. The source code can be downloaded at `http://www.iiia.csic.es/~ipinyol/sourceIJAR10.zip`.

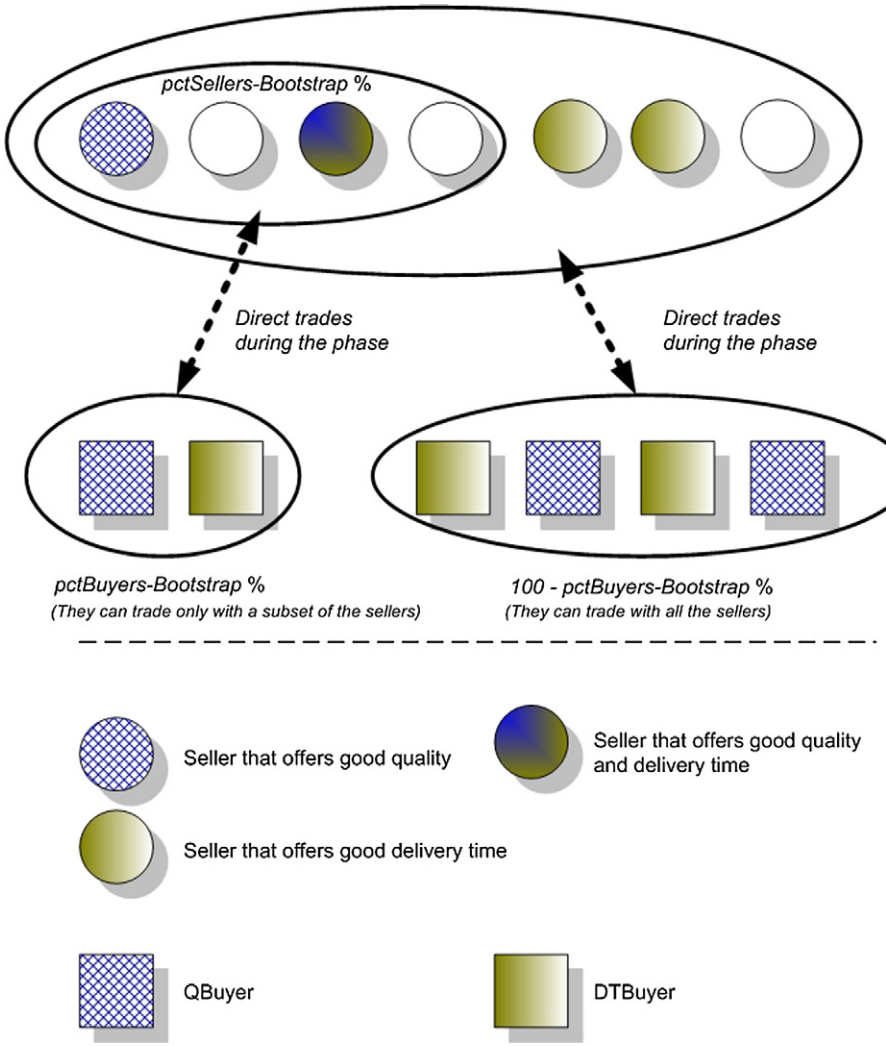

Fig. 5. A possible scenario during the bootstrap phase.

buyers can trade directly to any seller. Such special buyers will have to model the remaining sellers only using third-party information. Fig. 5 illustrates the scenario in this phase.

- *Experimental phase:* After the bootstrapping phase we introduce a single Q-buyer agent (our subject of study) that wants to model the behavior of a set of sellers (which correspond to a 100 - pctSellers-Bootstrap% of the sellers[7]) before trading with one of them. As discussed earlier, both the source and the recipient of the communication must have some knowledge before arguing, and because of that, our subject of study needs to go also through a bootstrap phase.

  An intuitive example that resembles into the structure of this phase is a situation in which a human buyer navigates through on-line forums starting new traces before making the decision of acquiring an expensive good (like a laptop, a car, etc.).

  Once the subject of study finishes the bootstrap, the simulation proceeds. As said before, it wants to buy a good, and for this, it needs to model the behavior of the unknown sellers. It receives a message from each buyer agent about each of the unknown sellers. Depending on the experimental condition, the studied agent will aggregate the communication to its reputation theory, or will argue and decide whether to accept or not the message. See Fig. 6 for an illustration of this phase.

The two experimental conditions are:

- **NO-ARG**: The studied agent does not use argumentation when receiving the messages. This means that the reputation model is the only mechanism to avoid information from bad informants. Agents use an extension of the Repage system [9] that contemplates an ontological dimension. In any case, the reputation model is able to detect bad informants

[7] We use the same *pctSellers-Bootstrap* value, but the set of sellers is not necessary the same as in the bootstrap phase.

comparing what they said with what they experienced. We recall here that we have two groups of buyer agents (QBuyer and DTBuyer) and that have different perspectives of what a *good* seller is.

- **ARG**: The studied agent and the source buyer agent (informant) engage in a dialog following the protocol described in the paper. Parameter β plays a crucial role in the experiments. It is easy to see that the higher the β parameter, the closer the performance results to be to NO-ARG condition, since when β is high enough, the argument is always accepted [10].

The main parameters that we manage in the simulations are:

Parameter	Description
#sellers	Number of sellers (40)
#buyers	Number of buyers (20)
pctGoodQuality	Percentage of sellers that offer good quality (25%)
pctGoodDTime	Percentage of sellers that offer good delivery time (25%)
pctQBuyers	Percentage of QBuyers. 100 - **pctQBuyers** are DTBuyer (20%, 50%, 80%)
pctBuyers-bootstrap	Percentage of buyers that during the bootstrap phase can only trade with a subset of randomly selected sellers that represent the percentage pctSellers-bootstrap
pctSellers-bootstrap	It also determines the percentage of sellers that during the experimental phase the studied agent does bootstrap with. Then, the rest of sellers are those that the studied agent must discover only through messages (20%).
turnsBootstrap	Turns in the bootstrap phase. We use such parameter to control the amount of initial information
β	Inconsistency budget (0)

For the simulations, agents use the L_{rep} language taking M as $\langle\{VB, B, N, G, VG\}, \leq_s\rangle$ (see Section 2). The performance of an execution computes how well the studied agent models the unknown sellers. We compare the best possible social evaluation as a seller (according to the parameters of the seller and the goals of the studied agent), with the real evaluation. For instance, given a seller who offers a bad quality and a very good delivery time, the best theoretically evaluation for an agent that is only interested in the quality dimension should be B (bad). In the case that our studied agent has evaluated such seller as G, the difference between both evaluations gives a measure of the achieved accuracy.

We use the difference function $\ominus$ defined over M, in which we consider a mapping $f : \{VB, B, N, G, VG\} \rightarrow [0, 4] \cap \mathbb{N}$, where $f(VB) = 0, f(B) = 1, f(N) = 2, f(N) = 3, f(N) = 4$, and define $\ominus(X, Y) = |f(X) - f(Y)|$. Then, 0 is the minimum difference, when both evaluations have the exact same value, and 4 is the maximum, when one evaluation is VG and the other VB.

We define the accuracy as the percentage of improvement with respect to the expected difference of two random evaluations. Given two random evaluations, their expected difference is exactly $\frac{40}{25} = 1.6$. The computation is summarized in the following table:

Difference	Possible values	Prob.	Expected
4	(VB, VG)	$2 \cdot \frac{1}{5} \cdot \frac{1}{5}$	0.32
3	$(VB, G), (B, VG)$	$2 \cdot 2 \cdot \frac{1}{5} \cdot \frac{1}{5}$	0.48
2	$(VB, N), (B, G), (N, VG)$	$2 \cdot 3 \cdot \frac{1}{5} \cdot \frac{1}{5}$	0.48
1	$(VB, B), (B, N), (N, G), (G, VG)$	$2 \cdot 4 \cdot \frac{1}{5} \cdot \frac{1}{5}$	0.32
		Total	1.6

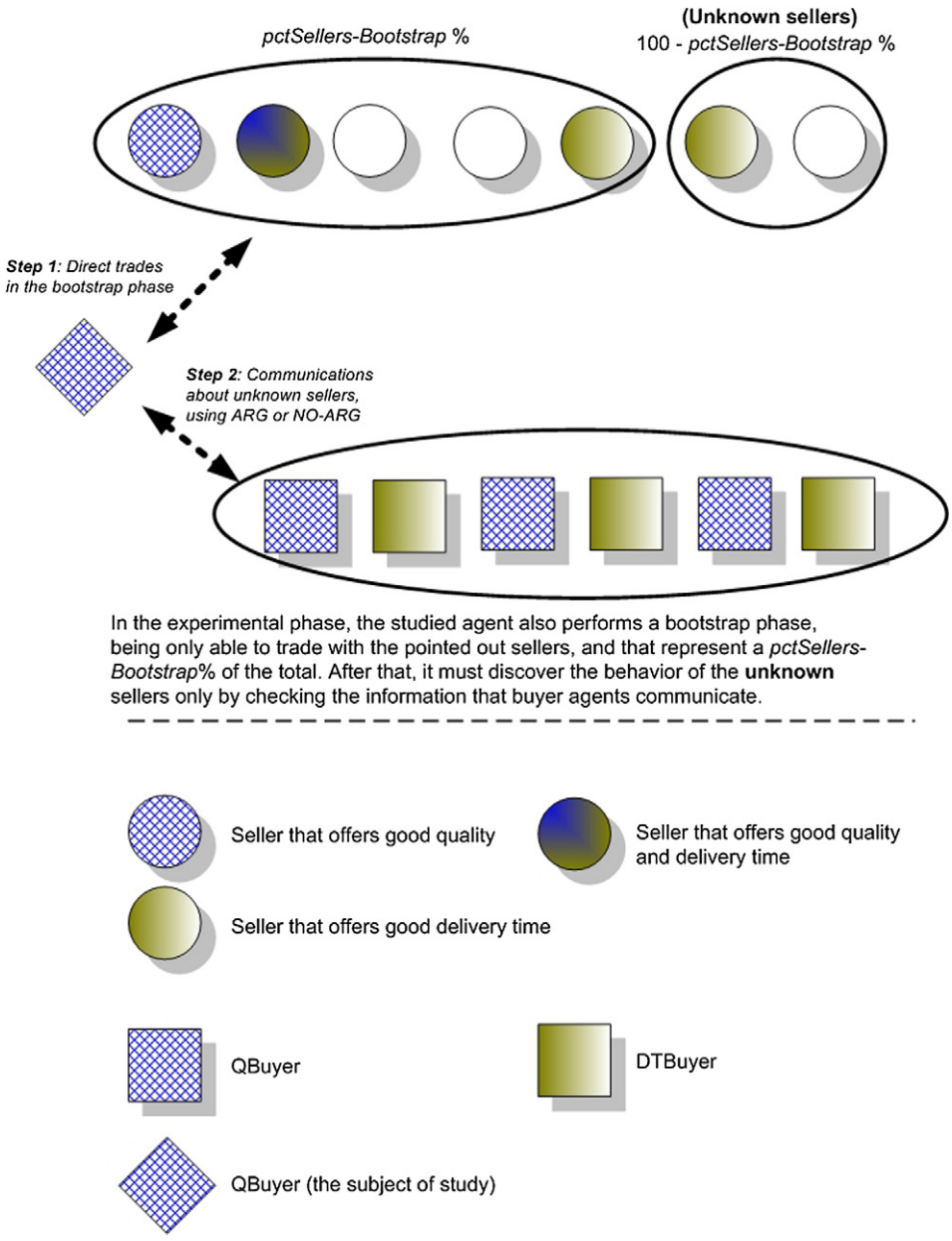

Fig. 6. A possible scenario during the experimental phase.

For instance, to compute the expectation of obtaining a difference of 2, we have to realize that there are only three situations in which this occurs: (VB, N), (B, G), (N, VG). Therefore, the probability of archiving a situation in which the difference is two is $3 \cdot \frac{1}{5} \cdot \frac{1}{5}$. Since we also consider the symmetric situation (so, (VB,N) and (N,VB)), the probability is in fact $2 \cdot 3 \cdot \frac{1}{5} \cdot \frac{1}{5} = 0.24$. Thus, the expectation value is $0.24 \cdot 2 = 0.48$.

One would expect that the reputation model improves such value, so, that the difference decreases to some extend from 1.6. For this, we calculate the average difference of all the unknown sellers and compute the percentage with respect to 1.6. For instance, an average difference of 0.5 improves 68.75% ($68.75 = (1.6 - 0.5) \cdot 100/1.6$), while an average difference of 0.3 improves 81.25% the random expected difference. We compare the experimental conditions ARG and NO-ARG using this measure.

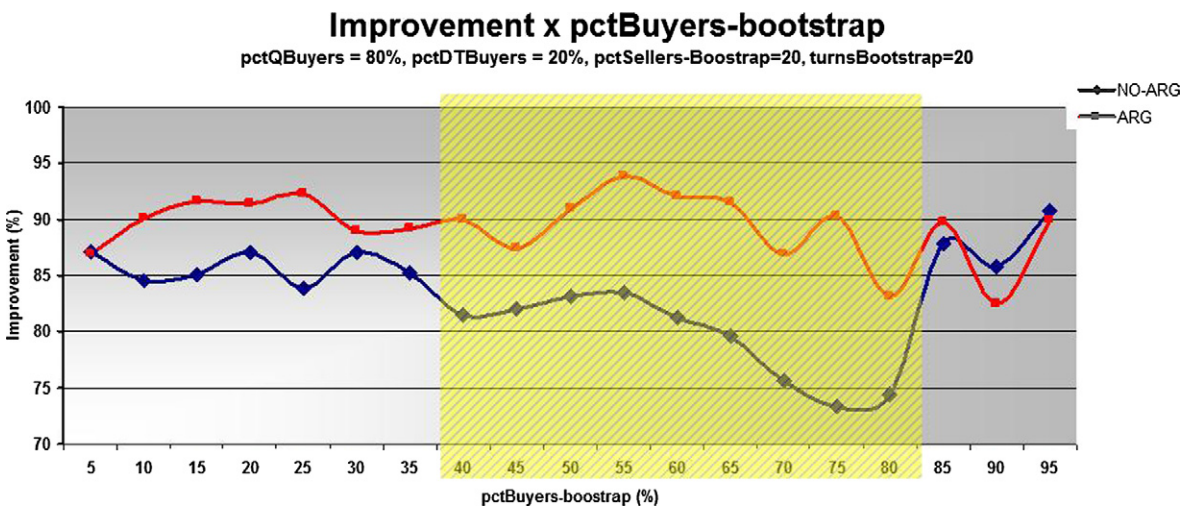

Fig. 7. The performance of both experimental conditions varying the *pctBuyers-Boostrap*, with pctQBuyers=80% and pctDTBuyers = 20%.

6.2. Simulation results

As said above, our main concern is to validate that the use of our argumentation mechanism improves significantly the accuracy of the agents in some conditions, and characterize them to some extend. Concretely, and in the terms used in the description of the experiment, the hypothesis is:

H: The experimental condition *ARG* achieves a higher improvement than the condition *NO-ARG*

We analyze such statement through the parameters *pctBuyers-Bootstrap* and *turnBootstrap*, because they model the amount of initial information that they have about the environment.

6.2.1. pctBuyers-bootstrap

The parameter models the number of buyers that cannot interact with all the sellers in the bootstrap phase, only with a subset of them. When the parameter is high it indicates that most of the buyers are not able to directly explore some sellers, and when it is low, that most of the buyers are able to explore all the sellers. This parameter is an indicator of how well the set of buyers is informed. We theorized that too little information as well as too much, can be critical in the use of argumentation. The simulation results are in tune with this idea.

Figs. 7, 8 and 9 show the performance of *ARG* and *NO-ARG* when varying *pctBuyers-Bootstrap* from 5% to 95% (setting *turnsBootstrap* to 20) with pctQBuyers=80%, pctQBuyers=50% and pctQBuyers=20% respectively . The results confirm the hypothesis for most of the points in the graph[8] (we highlight such intervals in the figures). It is interesting to observe that all three graphs show a range of *pctBuyers-Bootstrap* in which the hypothesis is always confirmed with a $p\_value \leq 0.01$. The following table summarizes them:

pctQBuyers	Intervals (%)
80% (fig. 7)	40–80
50% (fig. 8)	35–70
20% (fig. 9)	5–70

When *pctBuyers-Bootstrap* is higher than 80%, ARG does not improve significantly NO-ARG. Those are the cases where the lack of ground information makes the argumentation process useless. Also, when *pctBuyers-Bootstrap* is low, ARG does not necessary provide significant improvements over NO-ARG. This happens when the agents mostly have ground information, so the studied agent cannot reject any argument.

[8] We performed t-test statistical analysis to validate whether *ARG* significantly improves *NO-ARG* with a $p\_value \leq 0.01$. When this is the case, we say that H is confirmed. For the statistical analysis, each simulation is repeated 20 times. To give arguments in favor of assuming normality on the distributions, we applied the Jarque-Bera (JB) test for normality, and we could not reject the null hypothesis, which assumes that the distribution follows a normal distribution.

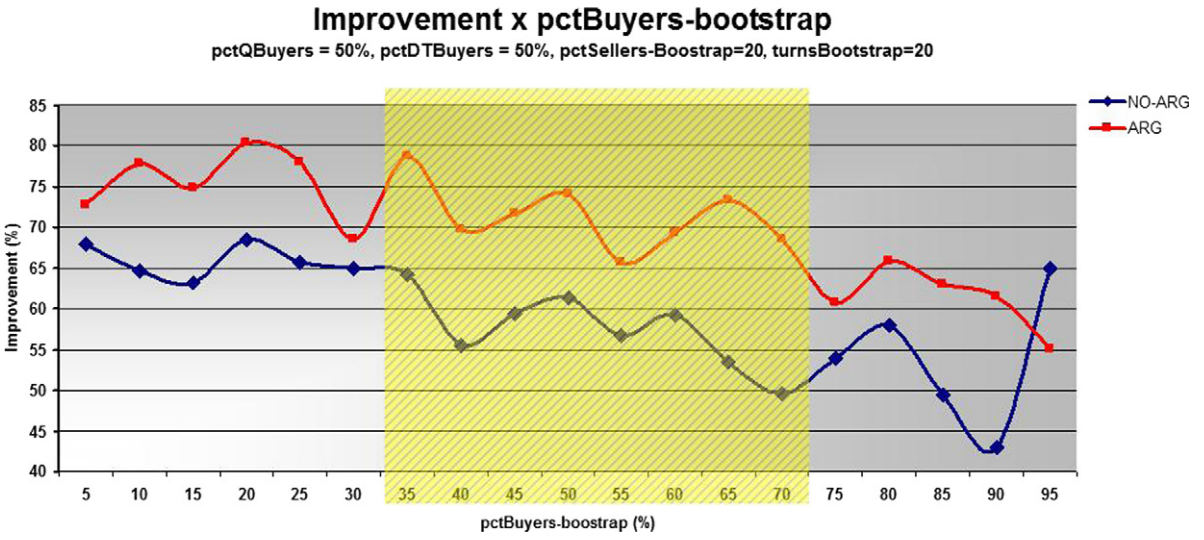

Fig. 8. The performance of both experimental conditions varying the *pctBuyersBoostrap*, with pctQBuyers=50% and pctDTBuyers = 50%.

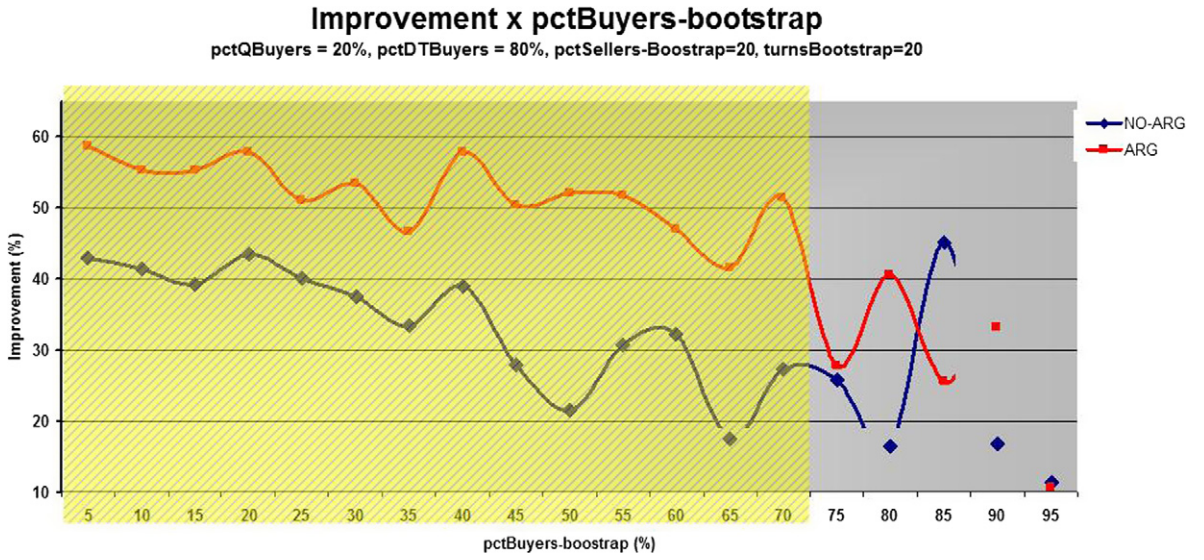

Fig. 9. The performance of both experimental conditions varying the *pctBuyers-Boostrap*, with pctQBuyers=20% and pctDTBuyers = 80%.

It is also interesting to observe that as *pctBuyers-Bootstrap* increases, the accuracy of both ARG and NO-ARG decreases a bit. This is because direct trades offer always a better way to discover sellers than just communications. Then, when *pctBuyers-Bootstrap* is high, less buyers can directly interact with all the sellers.

6.2.2. turnBootstrap

Related to the previous parameter, we want to study the amount of information that is needed to actually achieve an improvement by using argumentation. We set the parameter *pctBuyers-Bootstrap* to 75% and vary the turns that agents spend in the bootstrap phase. The higher *turnsBootstrap* is, the larger amount of information the agents will have about the sellers when the experimental phase starts.

Figs. 10, 11 and 12 illustrate the results with pctQBuyers=80%, pctQBuyers=50% and pctQBuyers=20% respectively. As expected, ARG does not perform better than NO-ARG until certain amount of data is managed by the agents. Fig. 12 is maybe the most illustrative situation. There it can be observed how from 10 turns on, ARG is always better than NO-ARG. This is an indicator of the amount of information needed to take advantage of argumentation. The following table summarizes the intervals where ARG significantly improves NO-ARG.[9]

[9] Not all points achieve a $p\_value < 0.01$. All of them though achieve $p\_value < 0.05$.

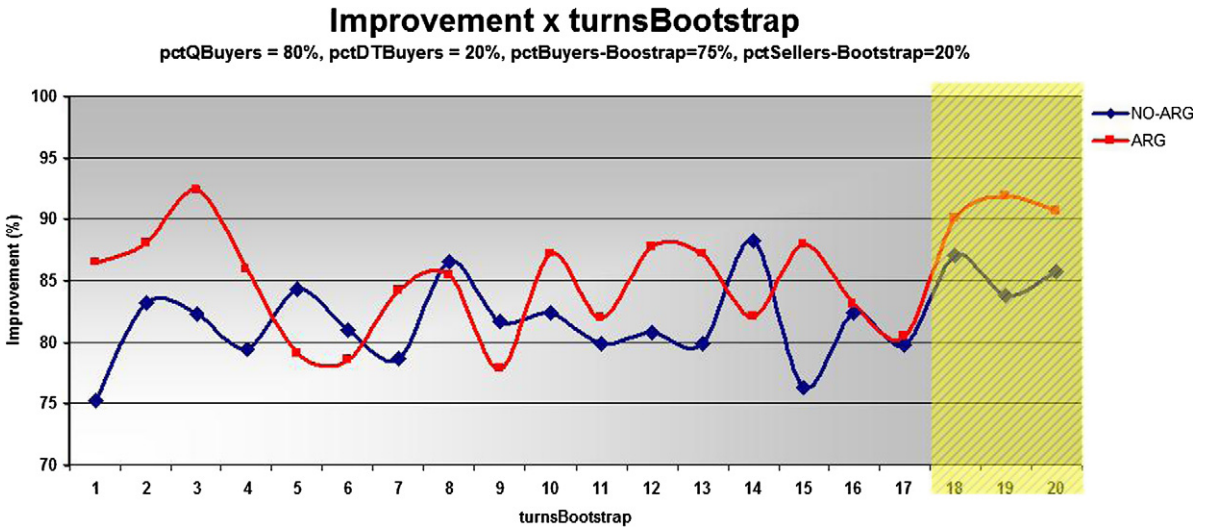

Fig. 10. The performance of both experimental conditions varying the *turnsBootstrap*, with *pctQBuyers*=80% and *pctDTBuyers* = 20%.

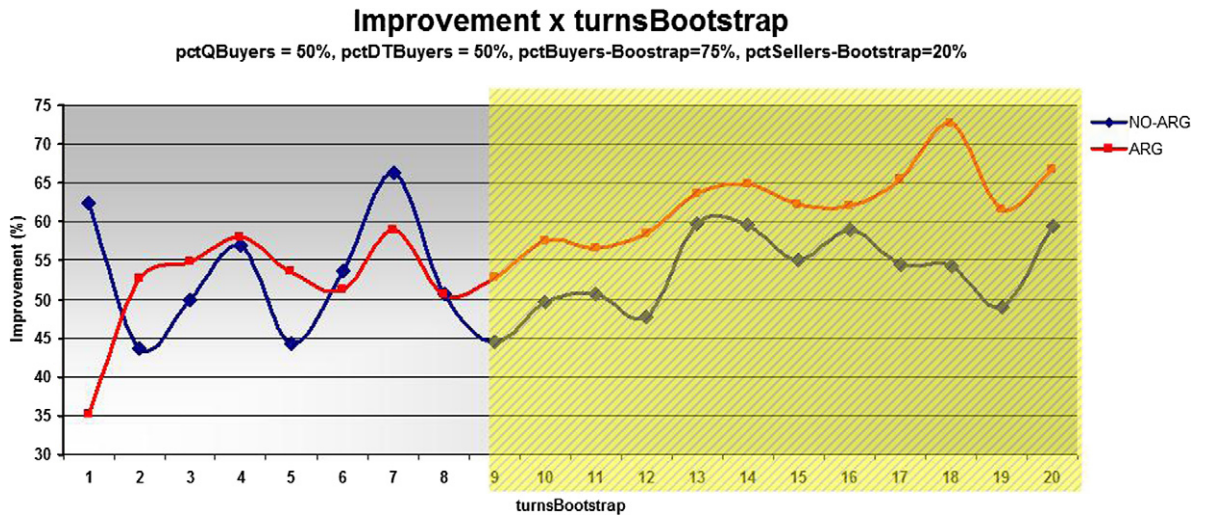

Fig. 11. The performance of both experimental conditions varying the *turnsBootstrap*, with *pctQBuyers*=50% and *pctDTBuyers* = 50%.

pctQBuyers	Intervals (turns)
80% (Fig. 10)	18–20
50% (Fig. 11)	10–20
20% (Fig. 12)	10–20

The intervals show the regions in which the difference is statistically significant with a $p\_value < 0.01$. However, some other points achieve a $p\_value < 0.05$, which in many cases would be enough to consider it a significant improvement.

The *pctQBuyers* shows an interesting behavior too. When it is 20% the improvement can be already appreciated in the turn 10, while when it is 50% and 80% the improvement is appreciated much later. We recall here that the studied agent is always a *QBuyer*, and then, when *pctQBuyers* is low, there are few *QBuyers*, so, few agents with the exact same goals. Therefore, the results confirm that when the percentage of *QBuyers* is low, few bootstrap turns are enough to encourage the use of argumentation. Notice that when *pctQBuyers* is high, the achieved improvement by NO-ARG is already high, while it is low when *pctQBuyers* is low. We can extrapolate here that when everybody has similar goals it is not worth arguing, while when it is not the case, argumentation can improve significantly the performance. The problem is that in real scenarios, it is hard to know the goals of the agents beforehand. Nevertheless, they could be learned by the agents.

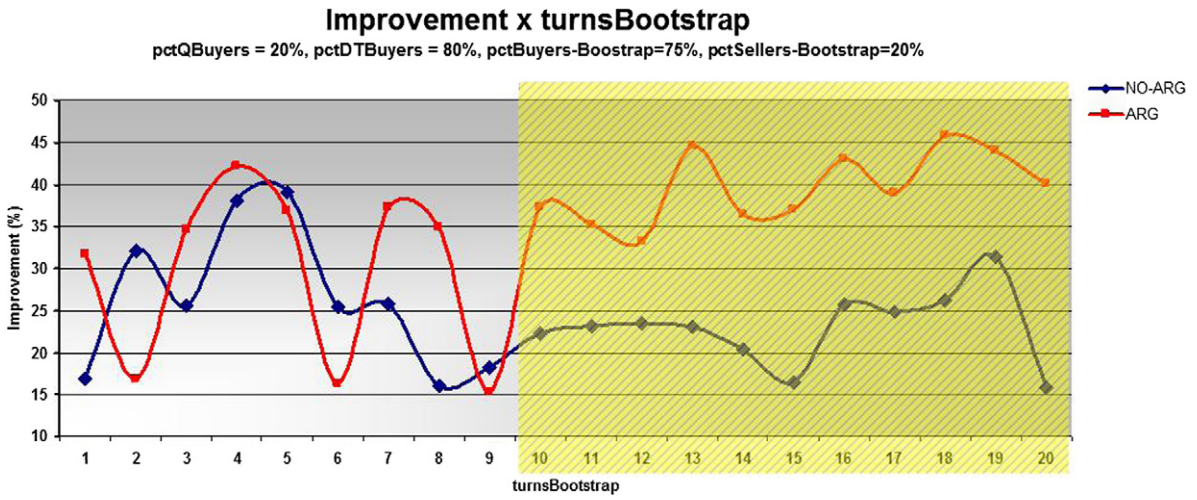

Fig. 12. The performance of both experimental conditions varying the *turnsBootstrap*, with *pctQBuyers*=20% and *pctDTBuyers* = 80%.

6.3. Discussion

We have performed a set of simulations to empirically validate the benefits of the argumentation-based protocol. We compare the accuracy obtained by agents using our argumentation protocol (ARG), and those not using it (NO-ARG). Our findings demonstrate that ARG significantly improves ($p\_value \leq 0.01$) the accuracy obtained by NO-ARG in most of the checked conditions. After the exploration of several parameters we conclude that when (1) there is a heterogeneity set of agents (not everybody has the same goals) and (2) agents do not base all their inferences in direct experiences, agents using argumentation achieve significantly a better accuracy that agents not using it.

When everybody has similar goals the gained accuracy using ARG may not be significant (1). The reason is that through argumentation, agents can reject information that they consider not reliable. On the contrary, when the goals are similar, the inclusion of the communications in the reputation theories does not produce many changes in the new deductions: the reputation mechanism by itself obtains very good accurate predictions that are difficult to be improved.

The analysis shows the importance of the bootstrapping phase, which models the fact that argumentation is useful when agents are endowed with some knowledge (2). Regarding this, the experiments reveal that certain number of bootstrapping turns are needed to make ARG better than NO-ARG. This situation is especially depicted in Fig. 12.

We want to remark that the presented simulations were performed independently for ARG and NO-ARG. It means that for each simulation, a bootstrap phase was executed and either ARG or NO-ARG where executed. We partially explored what happens when after the same bootstrap phase, we run ARG and NO-ARG. We observe that in most of the cases ARG performs better than NO-ARG. Preliminary results are illustrated in Fig. 13. We let for future work a more exhaustive exploration in this direction.

7. Related work

Many surveys exist in the literature regarding computational trust and reputation models. Some of them are based on on-line related systems [20–23] and others focused on peer-to-peer systems [24,25]. Some reviews tackle concrete aspects or functionalities like attack and defense techniques [26] or reputation management [27]. Others are more general [28–31]. All of them provide comprehensive definitions of models that use reliability measures calculated from the source of information.

Nevertheless, in this related work, we provide an overview of the work that takes advantage of the constituent elements of reputation-related concepts. For instance, models like [32,33] use *certified reputation*, in which the same target agent is able to justify its own reputation by presenting references (like reference letters when applying for a job). However, neither dialogs nor specific acceptability semantics is provided. Work presented in [34] explicitly uses argumentation techniques to handle recommendations. Its focus is bounded to peer-to-peer networks and recommendation systems. With a similar objective than the previous work the research published in [35] uses probabilistic argumentation techniques to calculate web of trusts applied to compute the strength of cryptographic keys. Thus, it relies on peer-to-peer networks and focuses on a very specific dimension of trust, instead of considering a set of components under which to argue. In [36] reputation values are justified by the history of interactions and social network analysis. In this approach, argumentation is used as a theoretical reasoning process, instead of a dialectical procedure.

The work presented in [37] contributes to the field of computational reputation models by analyzing different aggregation operations used by reputation models to better model the behavior of potential agents. This is in tune with the objective

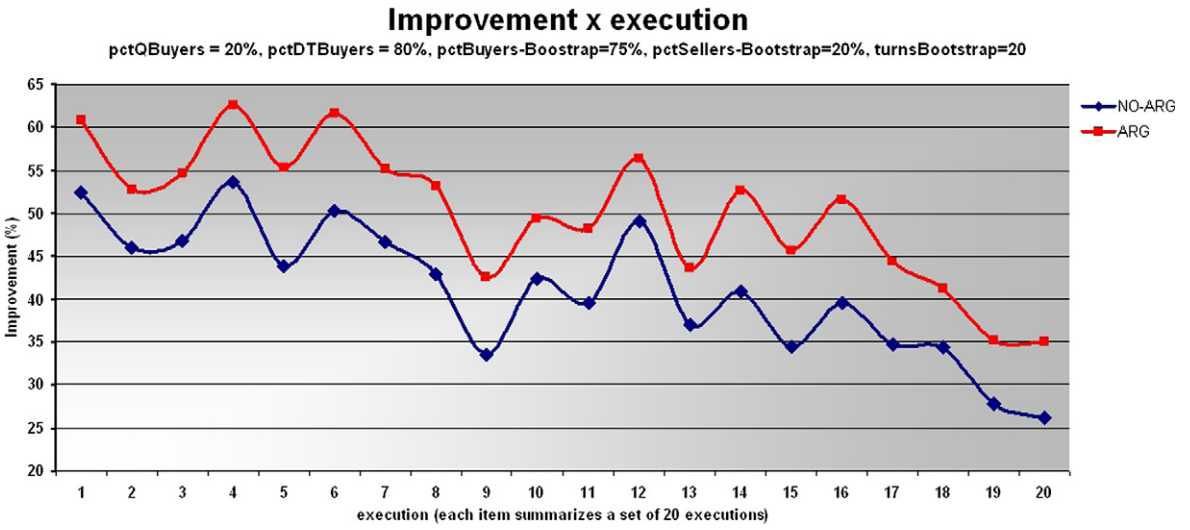

Fig. 13. The performance of both experimental conditions keeping the same bootstrap data, with fixed parameters. We illustrate how when keeping the same bootstrap data, argumentation always produces a better performance.

of our work, although in our case, we focus on a filter mechanism to prevent agents to aggregate useless communications, while the pointed out paper focuses on how such information must be aggregated inside the reputation model.

In [38], the authors analyze reputation-related concepts in terms of the internal elements used to infer them. However, it does not provide any formal protocol nor any acceptability semantics. More pragmatic approaches provide agent architectures for fuzzy argumentation on trust and reputation, but they lack formal definitions of acceptability [39].

A very recent work [40] presents a formal argumentation framework to argue about the trustworthiness of information sources using as a based the internal components of trust defined by Castelfranchi and Falcone [41]. However our approach offers an argumentation framework to deal with the components of social evaluations, which of course have a direct impact on the component called *core trust* by Castelfranchi and Falcone. Both approaches are complementary because they deal with different aspects of evaluations. While [40] deals with the components of Trust as defined by [41], we deal with the components of Social Evaluations, partially defined by Conte and Paolucci [4], and extended by Pinyol and Sabater-Mit [3]. Moreover, the pointed out paper remains at a theoretical level whilst we present an empirical validation that shows the benefits of using such kind of framework.

A promising research line that can be complementary to our approach comes from the solution of the trust alignment problem [42]. This approach suggests that with the exchange of ground elements to justify trust values (they consider only interactions, composed of contracts and fulfillments), it is possible to model other agents' inferences through inductive logic algorithms. This approach requires though a very stable social groups where agents can gather a lot of shared interactions and relatively simple reputation models.

From a more technical fashion, regarding the new argumentation-based protocol that we have developed, our work is related to [43]. One of the main features of our framework is the capability to handle both numerical and categorical values in the arguments, also used in the cited paper applied though to preference-based arguments. In a similar way, the work in [44] faces the problem of merging information from difference sources by using argumentation frameworks. We instantiate a similar framework to get in to the objective of our paper.

Finally, we do not want to forget the incursion of argumentation-based negotiation in the reputation and trust field. For instance, the work presented in [45] acknowledges the notion of trust as a multi-faced holistic construct, based on evaluable elements that can be used to argue and lead the decision making. We can say that the approaches are complementary. While our work focuses on the analysis of the internal elements of reputation-related components, which contributes to the field of computational reputation models, negotiation approaches try to integrate it in argumentation-based negotiation processes.

8. Conclusions

We have defined an argumentation-based protocol for the exchange of reputation-related information that allows agents to judge whether a given piece of information is reliable or not. We use argumentation techniques to give semantics to the protocol. The main characteristics of the system are:

- Only the recipient agent decides about the reliability of a communicated evaluation. This differs from other approaches in which the source agent attaches a reliability measure to the communicated social evaluation. This makes more difficult

for dishonest agents to intentionally send fraudulent information, because they must be aware of the knowledge of the recipient and justify the *lie* accordingly.

- It uses argumentation frameworks to give semantics to the dialogs. We exploit the L_{rep} language to completely define how arguments are constructed and influence one another. We instantiate a weighted abstract argumentation framework to define the acceptability semantics of a communicated social evaluation.
- It handles quantitative and qualitative graded information. One of the main characteristics of reputation information is that it is graded. Nowadays it is strange to find a model that provides crisp evaluations of the agents. For instance, an agent *A* may be *bad*, *very bad* or *very good* etc. as a car driver, and this has to be taken into account when arguing about evaluations.
- It permits dialogs between parties that use different reputation models. Even when we assume that agents use the same language to talk and reason about reputation information (L_{rep} language), we suppose that they can use different inference rules (different reputation models) without having to exchange the exact rules that each agent uses for the inferences.

We have made an important assumption: agents use the same language to *talk* about reputation concepts. This requires that the concepts described by the language have the same semantics for both agents. We allow though the use of different deduction rules to infer the predicates. In the case agents use different semantics they should engage first in a process of ontology alignment.

At the theoretical level, the next step regarding this work will be the inclusion of defeats among arguments. We plan to use the typology of ground elements to give strength to the arguments, independently of their attack relations. For instance, one may consider that arguments based on direct experiences are stronger than those based on communications.

At the simulation level, the objective of the experimental section was to provide empirical evidences that in some open multi-agent environments, the use of our argumentation-based protocol significantly improve the performance of the agents, meaning that they are able to better forecast the behavior of other agents in the interactions. It was not our intention to prove any property or usage of the inconsistency budget, which in our simulations is always set to 0 to maximize the difference with the non-argumentation setting. In the current experimental settings, when the inconsistency budget increases, the results of ARG and NO_ARG become similar. We let for future work an in-depth study of scenarios and situations where the strategic use of the inconsistency budget shows empirical evidences of better performance.

It is also important to remark that dialog games in dynamic contexts may be neither sound nor complete [46]. This implies that we can only ensure the correctness of the presented dialog protocol when the internal information that agents have does not change during the game. This constraint can be too strong in certain scenarios, specially when each step of the dialog can be considered a communication that may change the internal mental state of the recipient agents. This is a research line that should be explored and investigated in the future.

References

[1] M. Luck, P. McBurney, O. Shehory, S. Willmott, Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing), AgentLink, 2005.
[2] I. Pinyol, J. Sabater-Mir, An argumentation-based dialog for social evaluations exchange (short paper), in: Proceedings of the 19th European Conference on Artificial Intelligence (ECAI'10), Lisbon, Portugal., IOS Press Amsterdam, 2010, pp. 997–998.
[3] I. Pinyol, J. Sabater-Mir, G. Cuni, How to talk about reputation using a common ontology: From definition to implementation, in: Proceedings of the Ninth Workshop on Trust in Agent Societies, Hawaii, USA, 2007, pp. 90—101.
[4] R. Conte, M. Paolucci, Reputation in artificial societies: Social beliefs for social order, Kluwer Academic Publishers., 2002.
[5] M. Miceli, C. Castelfranchi, Human Cognition and Social Agent Technology, John Benjamins, 2000, Ch. The Role of Evaluation in Cognition and Social Interaction, pp. 225–259.
[6] J. Grant, S. Kraus, D. Perlis, A logic for characterizing multiple bounded agents, Autonomous Agents and Multi-Agent Systems 3 (4) (2000) 351–387.
[7] eBay, eBay, http://www.eBay.com (2002).
[8] A. Abdul-Rahman, S. Hailes, Supporting trust in virtual communities, in: Proceedings of the Hawaii's International Conference on Systems Sciences, Maui, Hawaii, 2000.
[9] J. Sabater-Mir, M. Paolucci, R. Conte, Repage: Reputation and image among limited autonomous partners, Journal of Artificial Societies and Social Simulation (JASSS) 9 (2). http://jasss.soc.surrey.ac.uk/9/2/3.html.
[10] P. Dunne, A. Hunter, P. McBurney, S. Parsons, M. Wooldridge, Inconsistency tolerance in weighted argument systems, in: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'09), Budapest, Hungary, 2009, pp. 851–858.
[11] C. Chesñevar, G. Simari, Modelling inference in argumentation through labeled deduction: Formalization and logical properties, Logica Universalis 1 (1) (2007) 93–124.
[12] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (2) (1995) 321–358.
[13] D. Walton, E. Krabbe, Commitement in Dialogue: Basic Concepts of Interpersonal Reasoning, State University of New York Press, 1995.
[14] L. Amgoud, N. Maudet, S. Parsons, Modelling dialogues using argumentation, in: Proceedings of the Fourth International Conference on MultiAgent Systems, 2000, pp. 31–38.
[15] P. Dunne, T. Bench-Capon, Two party immediate response disputes: Properties and efficiency, Artificial Intelligence 149 (2) (2003) 221–250.
[16] J. Sabater-Mir, Trust and reputation for agent societies, Ph.D. thesis, IIIA-CSIC, Barcelona, Spain, 2003.
[17] I. Pinyol, M. Paolucci, J. Sabater-Mir, R. Conte, Beyond accuracy. reputation for partner selection with lies and retaliation, in: Proceedings of the MABS'07. Hawaii, USA, vol. 5003 of LNCS, Springer, 2007, pp. 128–140.
[18] R.C.W. Quattrociocchi, M. Paolucci, Dealing with uncertainty: Simulating reputation in an ideal marketplace, in: Proceedings of the Eleventh Workshop on Trust in Agent Societies. Estoril, Portugal, 2008.
[19] M.P.W. Quattrociocchi, Reputation and uncertainty. a fairly optimistic society when cheating is total, in: First International Conference on Reputation: Theory and Technology (ICORE 09), Gargonza Italy, 2008, pp. 215–226.

[20] A. Jøsang, R. Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2) (2007) 618–644., emerging Issues in Collaborative Commerce.
[21] T. Grandison, M. Sloman, A survey of trust in internet applications, IEEE Communications Surveys and Tutorials 3 (4).
[22] D. Artz, Y. Gil, A survey of trust in computer science and the semantic web, Web Semantics: Science, Services and Agents on the World Wide Web 5 (2) (2007) 58–71 software Engineering and the Semantic Web.
[23] S. Grabner-Kräuter, E.A. Kaluscha, Empirical research in on-line trust: a review and critical assessment, International Journal of Human–Computer Studies 58 (6) (2003) 783–812.
[24] E. Koutrouli, A. Tsalgatidou, Reputation-based trust systems for p2p applications: Design issues and comparison framework, Trust and Privacy in Digital Business, LNCS, vol. 4083, Springer, 2006, pp. 152––161.
[25] G. Suryanarayana, R.N. Taylo, A survey of trust management and resource discovery technologies in peer-to-peer applications, ISR Technical Report UCI-ISR-04-6, University of California, 2004.
[26] K. Hoffman, D. Zage, C. Nita-Rotaru, A survey of attack and defense techniques for reputation systems, Tech. Rep. CSD TR 07-013, Purdue University, 2007.
[27] S. Ruohomaa, L. Kutvonen, E. Koutrouli, Reputation management survey, in: ARES '07: Proceedings of the The Second International Conference on Availability, Reliability and Security, IEEE Computer Society, Washington, DC, USA, 2007, pp. 103–111.
[28] I. Pinyol, J. Sabater-Mir, Computational trust and reputation models for open multi-agent systems: a review, Artificial Intelligence Review (2011) http://dx.doi.org/10.1007/s10462-011-9277-z.
[29] J. Sabater, C. Sierra, Review on computational trust and reputation models, Artificial Intelligence Review 24 (1) (2005) 33–60., doi:http://dx.doi.org/10.1007/s10462-004-0041-5.
[30] G. Lu, J. Lu, S. Yao, J. Yip, A review on computational trust models for multi-agent systems, in: International Conference on Internet Computing, 2007, pp. 325–331.
[31] L. Mui, A. Halberstadt, M. Mohtashemi, Notions of reputation in multi-agent systems: A review, in: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy, 2002, pp. 280—287.
[32] T. Huynh, N. Jennings, N. Shadbolt, An integrated trust and reputation model for open multi-agent systems, Journal of Autonomous Agents and MultiAgent Systems 2 (13) (2006) 119–154.
[33] E. Maximilien, M. Singh, Reputation and endorsement for web services, ACM SIGEcom Exchange 3 (1) (2002) 24–31.
[34] S. Heras, V.B.M. Navarro, V. Julian, Applying dialogue games to manage recommendation in social networks, in: ArgMAS 2009, 2009.
[35] R. Haenni, Using probabilistic argumentation for key validation in public-key cryptography, International Journal of Approximate Reasoning (IJAR) 38 (3) (2005) 355–376.
[36] J. Bentahar, J.C. Meyer, B. Moulin, Securing agent-oriented systems: An argumentation and reputation-based approach, ITNG, IEEE Computer Society, 2007, pp. 507–515.
[37] J. Sabater-Mir, M. Paolucci, On representation and aggregation of social evaluations in computational trust and reputation models, International Journal of Approximate Reasoning (IJAR) 46 (3) (2007) 458–483.
[38] I. Pinyol, J. Sabater-Mir, Arguing about reputation. the lrep language, 8th Annual International Workshop Engineering Societies in the Agents World, LNCS, vol. 4995, Springer, 2007, pp. 284–299.
[39] R. Stranders, M. de Weerdt, C. Witteveen, Fuzzy argumentation for trust, Proceedings of the Eighth Workshop on Computational Logic in Multi-Agent Systems (CLIMA-VIII), LNCS, vol. 5056, Springer, 2008, pp. 214–230.
[40] S. Villata,G. Boella, D.Gabbay, L. van der Torre, A socio-cognitivemodel of trust using argumentation theory, International Journal of Approximate Reasoning (IJAR) (2012) http://dx.doi.org/10.1016/j.ijar.2012.09.001.
[41] C. Castelfranchi, R. Falcone, Social trust, in: Proceedings of the First Workshop on Deception, Fraud and Trust in Agent Societies, Minneapolis, USA, 1998, pp. 35–49.
[42] A. Koster, J. Sabater-Mir, M. Schorlemmer, A formalization of trust alignment, in: 12th International Conference of the Catalan Association for Artificial Intelligence, Cardona, Catalonia, Spain, 2009.
[43] S. Kaci, L. van der Torre, Preference-based argumentation: Arguments supporting multiple values, International Journal of Approximate Reasoning (IJAR) 48 (3) (2008) 730–751.
[44] L. Amgoud, S. Kaci, An argumentation framework for merging conflicting knowledge bases, International Journal of Approximate Reasoning (IJAR) 45 (2) (2007) 321–340.
[45] M. Morge, An argumentation-based computational model of trust for negotiation, in: AISB'08, vol. 4, The Society for the Study of Artificial Intelligence and Simulation of Behavior, 2008, pp. 31–36.
[46] H. Prakken, Relating protocols for dynamic dispute with logics for defeasible argumentation, Synthese 127 (2000) 2001.