MS Dialogues: Persuading and getting persuaded

A model of social network debates that reconciles arguments and trust

Simone Gabbriellini and Paolo Torroni

DISI, University of Bologna
Viale del Risorgimento, 2
40136, Bologna - Italy
{simone.gabbriellini;paolo.torroni}@unibo.it

Abstract. We propose a new dialogue model to harmonise argumentation and trust. This can be fruitfully used in agent based social simulation experiments, to model a population of agents that interact within a social structure, exchange information by means of simulated discussions and possibly reach an agreement.

Keywords: Argumentation in multi-agent systems, argumentation dialogues, behavioural models, agent based social simulation, abstract argumentation, social networks,

1 Introduction

The formalization of dialogues is a major research topic in argumentation [7]. A particular type of formal dialogues studied in the literature are *persuasion dialogues*. According to Walton [50], in such dialogues the goal of each party is to persuade the other party to accept some designated proposition, using as premises only propositions that the other party has accepted as commitments. The elements of a persuasion dialogue are two participants, called the proponent P and the respondent R, and two propositions α and β (statements): α is P's thesis, while β is R's thesis. Such theses stand in a relation of opposition to each other. P's goal is to prove α from R's commitments, and R's goal is to prove his own thesis from P's commitments [51].

A characteristic of persuasion dialogues is that agents have a *goal*–to persuade the other agent–and *commitments*, which agents stand by, until proven wrong. Other types of dialogues have been defined as well, by Walton and Krabbe [49], and by other authors in more recent times. These include *information-seeking dialogues*, where one participant seeks the answer to some question from another participant who is believed to know the answer, *inquiry dialogues*, where two parties collaborate to answer some questions, *negotiation dialogues* that start from a conflict of interest and seek a reasonable settlement, *deliberation dialogues* aimed at deciding the best available course of action, and *eristic dialogues* where participants aim at verbally hitting out at the opponent.

Walton and Krabbe classify these types of dialogues along three dimensions: the initial situation, the participants' goals, and the goal of the dialogue. There are also subtypes, that capture the nuances between different initial conditions, and a number of other factors. Interestingly, the type of a dialogue is in fact a dynamic concept. For example, it may happen that the purpose of an ongoing dialogue dynamically shifts, alongside with the participants' goals, giving raise to what is known in the literature as *dialectical shifts* [26]. For example, a negotiation dialogue may take an eristic drift if the parties involved fail to find a reasonable settlement.

In the context of multi-agent systems, and in particular artificial societies, where (software) agents have clearly defined goals and interact by way of well defined protocols, the notion of argumentative dialogues has gathered considerable popularity, and several frameworks have been proposed for automating argumentative dialogues of various types (see [29] for a review of dialogue games for agent argumentation and [39] for a review of formal dialogue systems for persuasion).

Many such proposals use Dung's abstract argumentation [14] as a computational framework to model arguments and reason from them automatically, and define rules that agents must follow in order to produce meaningful dialogues of one type or another.

Following Hamblin [22] and McKenzie's dialectical system DC [27], it is common to define the steps of a dialogue in terms of *dialogue moves* associated to locutions such as *assert*, *accept*, *question*, *challenge*, *justify*, and so on, as it is done in formal dialogue frameworks such as those used in negotiation [41]. Each locution is defined in terms of its rationale, preconditions, intended effects (e.g., on the mind of the counterpart), and/or effects on the *state* of the dialogue, captured by a set of commitments (the *commitment store*), that define the epistemic position of each party in that particular dialogue [2,34]. Rules are typically defined, e.g., in the case of persuasion dialogues such as van Eemeren and Grootendorst's critical discussions [17], to ensure that both parties stick to the point and that they use commitments of the other party in their attempts to persuade the other party by means of rational argumentation.

This effort on formalising dialogues led to interesting results and applications in artificial societies, e.g., in the context of automated negotiation [40], and in the formalization of dialectical models of legal argument [38].

A recent domain of interest for (persuasion) dialogues is the social Web. An increasing number of organizations and businesses look at online dialogues and, more in general, at online debates, with growing attention, as these debates represent a valuable source of information about the sentiment of people about products, brands, policies, and so on. Online debates are a way for people belonging to a social network to exchange opinions and arguments, and can be influential in determining market trends and acceptance of new policies and regulations. Arguments in particular are the key to understand people's thinking and behaviour. Indeed, the debates occurring in online communities have important effects, e.g., in the diffusion of opinions and innovation in the information society. Here, argumentation technology can play a role in supporting this type of analysis. One possibility, which we advocate, is to model people's beliefs by way of argumentation frameworks, online debates by way of argumentative dialogues/debates, and then use agent-based social simulation techniques for describing and possibly forecasting the emergence of interesting properties, such as the likelihood that in the long run arguments will be more or less polarised, and that a given position or sentiment will win over other contrasting positions or sentiments.

However, the types of dialogue formalized in the argumentation literature seen so far are not entirely suitable for modelling dialogues occurring in online social networks. In social networking platforms we can surely find exchanges motivated by recognisable purposes such as persuasion, or information seeking, and there are some kinds of "rules of encounter" such as netiquettes. More frequently, however, the encounters and comments we find are serendipitous. They do not follow predetermined rules, nor can we easily identify motivating goals, preconditions and intended effects. Indeed, the term "bottom-up argumentation" has been coined to capture the emergence of arguments and opinions stemming for these types of exchanges [47,32].

To the best of our knowledge, in the state of the art there is no formal dialogue model that aims to capture the possible outcomes of natural dialogues or debates occurring in mainstream online social networks. There, we cannot take the perspective of informed professionals debating in institutional environment (such as in the legal domain), or software agents whose motivation can be described by a cost functions (such as in negotiations occurring in artificial societies). Rather, we ought to start from cognitive models of human communication, such as those proposed in experimental psychology.

An influential such model is Mercier & Sperber's argumentative theory of reasoning [30], whereby the emergence of reasoning is best understood within the framework of the evolution of human communication. The function of human reasoning is argumentative. Reasoning enables people to exchange arguments that, on the whole, make communication more reliable and hence more advantageous. In particular, for communication to be stable, it has to benefit both senders and receivers. To avoid being victims of misinformation, receivers must exercise some degree of *epistemic vigilance*.

Several psychological mechanisms may contribute to epistemic vigilance. The two most important ones are trust calibration and coherence setting. Some initial coherence checking occurs in the process of comprehension. When it uncovers some incoherence, an epistemically vigilant addressee must choose between two alternatives: either to reject communicated information, thus avoiding the risk of being misled, at the expense of possibly missing an opportunity to correct or update earlier beliefs, or to associate coherence checking and trust calibration, and allow for a fine-grained process of belief revision.

In particular, if a highly trusted individual tells us something that is incoherent with our previous beliefs, some revision is unavoidable. On the other hand, if a communicator wants to communicate a piece of information that the addressee is unlikely to accept on trust, she can produce arguments for her claims, and encourage the addressee to examine, evaluate, and accept these arguments. Reasoning contributes to the effectiveness and reliability of communication by allowing communicators to argue for their claim and by allowing addressees to assess these arguments.

This simple conceptual framework, where argumentation and trust together drive every communication between human actors, provides a very solid microfoundation for the dialogue model we intend to devise.

In this article, we propose an argumentation-based dialogue model whose aim is to capture the outcomes of interactions between humans in a social context. We call such interactions *"MS" dialogues* (after Mercier & Sperber). Agents involved in MS dialogues reason argumentatively, and implement epistemic vigilance using argumentation and trust together. We present the model (Sections 2 and 3), discuss some of its properties (Section 4), and its possible use in agent-based social simulation (Section 5). We conclude with Section 6, where we position our contribution with respect to related research on combining argumentation and trust and on the applications of argumentation technologies to online social platforms, and we discuss future work.

2 Agent model

Agents are characterised by their opinions and by the reasoning that leads them to maintain such opinions. Following Mercier & Sperber, agent reasoning is argumentative. In order to define a computational model, we adopt computational argumentation. In particular, we model opinions as abstract arguments in Dung-style argumentation frameworks. In the remainder of the paper, we will use the terms opinion, argument, and belief interchangeably.

In computational abstract argumentation, as defined by Dung [14], an *Argumentation Framework* (AF) is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a set of atomic arguments and $\mathcal{R}$ is a binary *attacks* relation over arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$, with $\alpha \rightarrow \beta \in \mathcal{R}$ interpreted as "argument α attacks argument β." Sets of "justified" arguments can be described by various extension-based semantics [4]. In particular, an extension-based semantics identifies a number of subsets of $\mathcal{A}$ that all together represent a coherent set of beliefs.

Argumentation semantics may be conflict-tolerant, whereby arguments in the same extension may attack each other [3], but most proposals require that extensions be *conflict-free*, i.e., no two arguments in the same extension may attack each other. In our model, we use conflict-free semantics. Therefore if an extension contains an argument α, either $\mathcal{R}$ does not contain any attack to α, or else α is defended, from all incoming attacks, by arguments in the extension.

Some well-known semantics defined by Dung are called *admissible*, *preferred*, and *complete* semantics. In particular, let S be a set of arguments, $S \subseteq \mathcal{A}$,

- S is *conflict-free* if $\forall \alpha, \beta \in S, \alpha \rightarrow \beta \notin \mathcal{R}$;
- an argument $\alpha \in S$ is *acceptable* w.r.t S if $\forall \beta \in \mathcal{A}$ s.t. $\beta \rightarrow \alpha \in \mathcal{R}$, $\exists \gamma \in S$ s.t. $\gamma \rightarrow \beta \in \mathcal{R}$;

- S is an *admissible* extension if S is conflict-free and all its arguments are acceptable w.r.t. S;
- S is a *preferred* extension if it is a maximal admissible set, w.r.t. set inclusion;
- S is a *complete* extension if S is admissible and $\forall\alpha \in \mathcal{A} \setminus S$, $\exists\beta \in S$ s.t. $\beta \rightarrow \alpha \in \mathcal{R}$.

Other extension-based semantics have been defined, such as the *grounded*, *stable*, *semistable*, and *ideal* semantics [4].

Not all the semantics guarantee that there is always an extension. In fact, in general a semantics may define, for a given AF, one or more extensions, or none at all. For example, according to many semantics, such as the *complete* semantics [14] and the *ideal* semantics [15], an argumentation framework such as $\langle\{a, b\}, \{a \rightarrow b, b \rightarrow a\}\rangle$ admits at least two extension: $\{a\}$ and $\{b\}$. Other semantics, such as the *grounded* semantics [14], require that there is only one extension. According to the grounded semantics, the only extension for the above AF is the empty set.

We do not commit to one semantics in particular, but, for generality, we assume that each agent may admit more than one extension at a time. We denote by $\mathcal{E}$ the set of all extensions of an agent's AF, defined by a given semantics.

The set of arguments $\mathcal{A}$ is the same for all agents. The differences between agents are in the attacks relations, $\mathcal{R}$. The reason behind this modelling choice is that in order for two agents to communicate, they must share a common language, and such language is made by the arguments and the attacks. This does not imply a loss in generality, as we could think of $\mathcal{A}$ as the set of all possible arguments of all agents, and use self-attacks to "mask" selected arguments from certain agents, if needed (no conflict-free semantics would include a self-attacking argument in any extension). If a self-attack relation is removed as a result of a dialogical exchange, a new argument may be included in some extension. This models the possibility that agents become aware of new arguments as information is shared between them.

Moreover, different people may have different understandings of the same arguments. For example, two conflicting arguments may be:

> *(a) Sugar mills produce as much as windmills produce, and at half the cost. Therefore, sugar mills are preferable to windmills*
> *(b) Recent studies show that windmills are much more energy-efficient than sugar mills. Therefore, windmills are preferable to sugar mills.*

Simplifying, we can define a as an argument supporting sugar mills against windmills, and b as one in supporting windmills against sugar mills. So let us consider, for the sake of illustration, a possible natural dialogue between two people, say, *Alice*, initially supporting a, and *Bob*, initially supporting b.

We could think that, initially, Alice has a strong argument supporting sugar mills against windmills (*sugar mills produce as much as windmills produce, and at half the cost. Therefore, sugar mills are preferable to windmills*), and a weak argument for the contrary (*windmills are preferable to sugar mills*). This is a *weak* argument because it does not really rely on any sequence of premises that bring to the conclusion. On the other hand, Bob has a weak argument supporting sugar mills against windmills (*sugar mills are preferable to windmills*), and a strong argument for the contrary (*recent studies show that windmills are much more energy-efficient than sugar mills. Therefore, windmills are preferable to sugar mills.*).

In fact, Alice's argument supporting windmills against sugar mills is so weak that Alice does not consider it to be worth undermining her argument supporting sugar mills against windmills. Likewise, Bob does not consider his argument against windmills to be strong enough to represent an attack against his argument supporting windmills.

In summary, before communicating with each other, Alice and Bob could be described by $AF_A = \langle \mathcal{A}, \mathcal{R}_A \rangle$ and $AF_B = \langle \mathcal{A}, \mathcal{R}_B \rangle$, where $\mathcal{A} = \{a, b\}$, $\mathcal{R}_A = \{a \to b\}$, and $\mathcal{R}_B = \{b \to a\}$, and A's only complete extension is a, whereas Bob's complete extension is b.

Then new elements may emerge during the dialogue, that change the relative strength of a and b. For example, Alice may get convinced by Bob that the case for windmills is as strong as, or even stronger than, the case for sugar mills. Accordingly, as a result of their dialogue, Alice's attack relation will change into $\mathcal{R}'_A = \{a \to b, b \to a\}$ or $\mathcal{R}''_A = \{b \to a\}$, and Alice's complete extensions will be $\mathcal{E}'_A = \{\{a\}, \{b\}\}$ or $\mathcal{E}''_A = \{\{b\}\}$.

Since we adopt an abstract argumentation approach, we do not describe how arguments are built. In particular, we do not analyse the elements constituting a and b and their evolution as the dialogue between Alice and Bob unfolds. However, we do model the shifts in the relations between a and b, by letting agents modify their sets of attacks.

Other modelling choices are indeed possible. In particular, we could consider two natural arguments to be different unless they share the same premises, conclusion, and logical inferences. For example, we could say that *windmills are preferable to sugar mills* isn't the same argument as *recent studies show that windmills are much more energy-efficient than sugar mills. Therefore, windmills are preferable to sugar mills.*

We could also assume that all agents have the same attack relation, based on a shared notion of conflict, whereas they may have different defeat relations. For example, if an agent thinks that it is more important to save energy than to save money, then they will find that the argument for windmills defeats the argument for sugar mills. Then the effect of dialogues would be that agents may end up with different defeat relations. Value-based argumentation frameworks [6] or extended argumentation frameworks [31] could accommodate such a model.

Our simpler and in a sense more abstract modelling choice, where opinions and arguments supporting them coincide, and there is a dynamic attacks relation, is motivated by our main research goal: to develop a simple model of bottom-up dialogues that take place in social networks, and study the propagation of opinions and phenomena such as polarisation in social networks (see Section 5). A comprehensive empirical validation of our model should tell us if our model can indeed capture the outcomes of discussions between humans, and whether it should be changed or extended.

Agents represent human actors situated in social networks, and they interact with one another by way of MS dialogues. For example, an MS dialogue could be the abstract representation of a sequence of posts (comments, *tweets* [19], etc.), in a given social networking platform. In general, more than two agents may be involved in a stream of posts. However, in this work we focus on two-party MS dialogues, leaving multi-party interactions for future extensions.

When agents communicate with one another, they exchange arguments and relations between arguments. Therefore a *post* can either be an argument α, or an attack $\alpha \rightarrow \beta$.

To reflect Mercier & Sperber's theory, in devising the MS dialogue model we followed these general guidelines:

- agents use argumentative reasoning, in order to establish the coherence of the information contained in the posts, against their own beliefs
- the information that an agent (author) contributes to a dialogue (i.e., an argument or an attack), is coherent with her own beliefs, i.e., it is included in one of the agent's extensions;
- agents evaluates posts using mechanisms for epistemic vigilance, based on argumentation and trust;
- the trust of an agent towards another may change dynamically as the dialogue evolves;
- if a post is incoherent with the recipient agent's beliefs, and the recipient trusts the post's author, she should actuate some form of belief revision in order to assimilate the new beliefs, while maintaining coherence;
- if a post is incoherent with the recipient agent's beliefs, and the recipient does *not* trust the post's author, she may either engage in an MS dialogue with the post's author, by producing arguments against the post, or simply ignore the post. In turn, the author can produce arguments for her claims, and encourage the recipient to examine, evaluate, and accept these arguments.

During a dialog, an agent is constantly assessing whether (a) the new information is coherent with her beliefs, (b) new arguments suffice to accept the new piece of information, and (c) in case of new incoherent information that requires revising beliefs, whether the counterpart is to be trusted or not.

We assume that agents rely on a trust model. Arguably a realistic model of trust would need to be a dynamic measure that takes into account the *nature of social ties*, and the *authoritativeness*, *expertise* and *social status* of the interlocutor [44]. However, our dialogue model is orthogonal to the trust model, and different trust models can be accommodated.

3 MS dialogues

In this section we explain how MS dialogues unfold.

Table 1. Notation

$\mathcal{A}$	a set of arguments, each identified by a letter of the alphabet, e.g., $\mathcal{A} = \{a, b, c, d\}$. We assume that $\mathcal{A}$ is the same for all agents
$\mathcal{R}$	a set of binary *attacks* relations between arguments, e.g., $\mathcal{R} = \{a \rightarrow b, c \rightarrow c\}$, where $a, b, c \in \mathcal{A}$
AF	AF=$\langle\mathcal{A}, \mathcal{R}\rangle$: an argumentation framework
ϵ	$\epsilon \subseteq \mathcal{A}$: an extension
$\mathcal{E}$	$\mathcal{E}$=$\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$: a set of n extensions
$\bigcup_{\mathcal{E}}$, $\bigcup_{\mathcal{E}}^{i}$	$\bigcup_{\mathcal{E}}$=$\epsilon_1 \cup \epsilon_2 \cup \cdots \cup \epsilon_n, \forall \epsilon_i \in \mathcal{E}$: the set of all arguments in each member of $\mathcal{E}$ (of an agent i).
A, B	agent names
$\mathcal{D}$	a dialogue between two agents. A dialogue is about a *subject*
$\sigma(\mathcal{D})$	the subject of $\mathcal{D}$
$\Delta_{in}, \Delta_{out}$	in/out *narratives*, i.e., all the *attacks* that an agent received/sent since $\mathcal{D}$ started, e.g.: $\Delta_{in} = \{a \rightarrow b, c \rightarrow b\}$

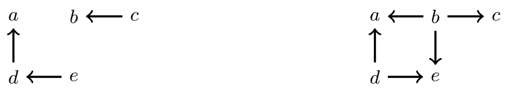
Let us stress once more that an MS dialogue is the abstract representation of a possibly complex exchange. The purpose of our model is not to define rigid rules for exchanging arguments in social networks, which would make little sense here, but rather, to define the possible outcomes of a dialogue, and the effect of a dialogue on the state of the participants. In particular, a dialogue can result in one agent or another changing her mind, or in both agents changing their mind (which does not necessarily imply an agreement), or it may have no effect at all on the participants' opinions.

Table 1 is there to guide the reader through our notation.

A dialogue $\mathcal{D}$ occurs between two agents A and B, about a subject $\sigma(\mathcal{D})$. The agent who starts the dialogue, say A, is called the *initiator*. Since he is starting a dialogue, we assume that the initiator has a claim to make, i.e., that he has at least one, non-empty, extension. In particular, if A is the initiator, $\mathcal{E}_A$ is the set of all extensions computed from A's AF, and $\bigcup_{\mathcal{E}}^{A}$ is the set of all arguments in all of A's extensions, then $\bigcup_{\mathcal{E}}^{A} \neq \emptyset$. The subject $\sigma(\mathcal{D})$ is an argument, belonging to one of the initiator's extensions: $\sigma(\mathcal{D}) \in \bigcup_{\mathcal{E}}^{A}$.

Following a shared practice in the literature, we label contributions using *locution* identifiers: *initiate*, *attack*, *rebut*, *ok* (termination with agreement), and *sorry* (termination with disagreement). The notation "**utter** $\underline{\text{initiate}}(\mathcal{D})$" indicates that A initiates an MS dialogue about $\sigma(\mathcal{D})$. In other words, $\underline{\text{initiate}}(\mathcal{D})$ is an "invitation to discuss" from A to another agent, say B, which sparks a dialogue $\mathcal{D}$ between A and B.

Let us consider two agents A and B defined by the argumentation frameworks shown in Fig. 1. For the sake of illustration, from now on we will stick to the *complete* semantics. A's only complete extension is $\{a, c, e\}$, whereas B's only complete extension is $\{b, d\}$. Thus A can initiate a dialogue about a, or c, or e. B can instead initiate a dialogue about b, or about d.



(a) AF_A (b) AF_B

Fig. 1. Sample argumentation frameworks.

Once an argument has been thrown on the table, MS dialogues proceed by the two agents attacking each other's claims and justifying their own claims. To do so they take turns, and communicate attacks to each other. We keep track of the attacks that each agent has produced (Δ_{out}) and received (Δ_{in}) since the dialogue started. Δ_{out} and Δ_{in} are initially empty narratives.

Whenever B is addressed by A with $\underline{\text{initiate}}(\mathcal{D})$, B behaves according to Algorithm 1. Notice that Algorithm 1, as well as all other algorithms below, are written from the perspective of an agent B reacting to some input coming from an agent A.

By Algorithm 1, if $\sigma(\mathcal{D})$ is coherent with B's AF, the dialogue has no reason to continue: A and B agree on the subject and the dialogue ends. If, instead, $\sigma(\mathcal{D})$ is incoherent with B's AF, i.e., if $\sigma(\mathcal{D})$ is not included in any of B's extensions, that means that $\sigma(\mathcal{D})$ supports an opinion that B is not currently embracing. In the case of complete semantics, $\sigma(\mathcal{D})$ actually conflicts with B's opinions. Thus, B will have to decide whether to accept this new argument or not.

To this end, B will exercise a degree of epistemic vigilance. In particular, if B *trusts* A, B will believe what A says, and revise her own beliefs (i.e., her argumentation framework) in order to accommodate A's argument $\sigma(\mathcal{D})$ in at least one of her extensions.[1] If instead B does not trust A, B will produce an argument α against $\sigma(\mathcal{D})$ ($\underline{\text{attack}}(\alpha \rightarrow \sigma(\mathcal{D}))$) and wait for a reaction from A. Producing such an attack amounts to showing A reasons not to believe $\sigma(\mathcal{D})$.

Now it is A's turn to produce arguments for her claims, and encourage B to examine, evaluate, and accept these arguments.

Let β be $\sigma(\mathcal{D})$. By Algorithm 2, to react to an attack $\alpha \rightarrow \beta$, an agent (called B in the description of the algorithm, but it may be either the initiator or the interlocutor in different phases of the dialogue) has different options:

- if he is aware of $\alpha \rightarrow \beta$, then, if he can attack α with a fresh argument, he will do it (**utter** $\underline{\text{attack}}(\gamma \rightarrow \alpha)$, line 7), otherwise he will attack something his interlocutor said in the past, again, using a fresh argument (**utter** $\underline{\text{attack}}(\gamma \rightarrow \alpha')$, line 12);
- if he is unaware of $\alpha \rightarrow \beta$, and he trusts his interlocutor, he will update his own *attacks* relation to include $\alpha \rightarrow \beta$ (lines 19-21), which may result in he actually changing opinion and agreeing with his interlocutor (line 24), or may not, in which case he will challenge (line 28 calls Algorithm 2 again, but this time the agent is aware of $\alpha \rightarrow \beta$ and Algorithm 2 will end up in the previous case);

[1] We elaborate on belief revision at the end of this section.

Algorithm 1 React to initiate($\mathcal{D}$)

```
Require: received initiate($\mathcal{D}$) from $A$
 1: $\Delta_{in} \leftarrow \emptyset, \Delta_{out} \leftarrow \emptyset$
 2: $\mathcal{E} \leftarrow$ set of all extensions computed from $AF$
 3: if $\sigma(\mathcal{D}) \in \bigcup_{\mathcal{E}}$ then
 4:   utter ok {there is agreement, the dialogue ends}
 5: else
 6:   if $B$ trusts $A$ then
 7:     revise $AF$, in order to achieve $\sigma(\mathcal{D}) \in \bigcup_{\mathcal{E}'}$
 8:     {$\mathcal{E}'$ is the set of extensions in the revised $AF'$}
 9:     utter ok {the dialogue ends with agreement}
10:   else
11:     {$A$ is not trusted: the dialogue continues}
12:     repeat
13:       pick $\alpha \rightarrow \sigma(\mathcal{D}) \in \mathcal{R} \setminus \Delta_{out}$, such that $\alpha \in \bigcup_{\mathcal{E}}$
14:       utter attack($\alpha \rightarrow \sigma(\mathcal{D})$)
15:     until $\nexists \alpha \rightarrow \sigma(\mathcal{D}) \in \mathcal{R} \setminus \Delta_{out}$, such that $\alpha \in \bigcup_{\mathcal{E}}$ or $\mathcal{D}$ has ended
16:   end if
17: end if
```

– if he is unaware of $\alpha \rightarrow \beta$, and he does not trust his interlocutor, he will deny $\alpha \rightarrow \beta$ to his interlocutor (rebut($\neg(\alpha \rightarrow \beta)$), line 31). The interlocutor will then continue the dialogue according to Algorithm 3.

Algorithm 3 defines how an agent reacts to a denial of an attack $\alpha \rightarrow \beta$. Again the usual mechanisms of epistemic vigilance play their role. We will not comment on Algorithm 3 in detail, as it is similar to Algorithm 2.

An MS dialogue $\mathcal{D}$ may end in different ways:

– A and B agree on $\sigma(\mathcal{D})$ from the start (see Algorithm 1: **utter** ok, line 4). No revision of beliefs is needed;
– either A or B change her mind about $\sigma(\mathcal{D})$ (see Algorithm 2 line 23 and Algorithm 3: **utter** ok, line 8). A revision of beliefs is needed on A or B's side;
– A and B disagree on $\sigma(\mathcal{D})$ (see Algorithm 3: **utter** sorry, line 26), in spite of possible revisions of beliefs that may have occurred along $\mathcal{D}$.

These procedures define the general framework of how MS dialogues unfold. For the sake of generality, several choice points are left open. Mainly, we do not commit to any specific argumentation semantics, we do not define how agents should revise their beliefs, and we do not specify how trust is formed.

In fact, our implementation of the model (see Section 5) accommodates not just one, but several well known extension-based argumentation semantics: *admissible*, *grounded*, *complete*, *preferred*, *ideal*, *stable*, and *semi-stable*. The user can choose a different semantics at each simulation from a drop-down menu.

As far as belief revision, we implemented one particular form of revision, which models the outcome of a possible dialogue between A and B, where A

Algorithm 2 React to attack($\alpha \to \beta$)

```
Require: received attack($\alpha \to \beta$) from $A$
$\Delta_{in} \leftarrow \Delta_{in} \cup \alpha \to \beta$
$\mathcal{E} \leftarrow$ set of all extensions computed from $AF$
if $\alpha \to \beta \in \mathcal{R}$ then
  {the agent is aware of the relation $\alpha \to \beta$}
  if $\exists \gamma \to \alpha \in \mathcal{R} \setminus \Delta_{out}$ such that $\gamma \in \bigcup_{\mathcal{E}}$ then
    {there is another direct attack against $\alpha$}
    utter attack($\gamma \to \alpha$)
    $\Delta_{out} \leftarrow \Delta_{out} \cup \gamma \to \alpha$
  else
    {attack something A said in the past}
    pick $\gamma \to \alpha' \in \mathcal{R} \setminus \Delta_{out}$ such that $\alpha' \to \beta' \in \Delta_{in}$, and $\gamma \in \bigcup_{\mathcal{E}}$
    utter attack($\gamma \to \alpha'$)
    $\Delta_{out} \leftarrow \Delta_{out} \cup \gamma \to \alpha'$
  end if
else
  {the agent is unaware of the relation $\alpha \to \beta$}
  if $B$ trusts $A$ then
    {believe what A says and revise own $AF$}
    add $\alpha \to \beta$ to $\mathcal{R}$ if $\alpha \to \beta \in \mathcal{R}_A$
    add $\beta \to \alpha$ to $\mathcal{R}$ if $\beta \to \alpha \in \mathcal{R}_A$
    remove $\beta \to \alpha$ from $\mathcal{R}$ if $\beta \to \alpha \notin \mathcal{R}_A$
    compute $\mathcal{E}'$ from the revised $AF$
    if $\sigma(\mathcal{D}) \in \bigcup_{\mathcal{E}} \setminus \bigcup_{\mathcal{E}'} \cup \bigcup_{\mathcal{E}'} \setminus \bigcup_{\mathcal{E}}$ then
      {the agent changed her mind about the subject: $\mathcal{D}$ ends with agreement}
      utter ok
    else
      {admitting $\alpha \to \beta$ does not change $B$'s mind about $\sigma(\mathcal{D})$: $\mathcal{D}$ continues}
      react to attack($\alpha \to \beta$)
    end if
  else
    utter rebut($\neg(\alpha \to \beta)$)
  end if
end if
```

learns some bits of knowledge from B in order to be able to embrace B's argument. Within Algorithm 1, in order to revise her beliefs, an agent progressively learns the interlocutor's *attacks* relations, which are assimilated into her own AF, and possibly deletes some of her own *attacks* relations, until she admits at least one extension that includes $\sigma(\mathcal{D})$. This is done in a conservative way, i.e., an attack can be added from the revising agent's AF only if it is in the interlocutor's AF, and it can be removed only if it is not in the interlocutor's AF. Within Algorithms 2 and 3, the interlocutor's *attacks* relations between α and β are assimilated into the revising agent's AF (i.e., whatever attacks between α and β are defined in the interlocutor's AF will be added into the revising agent's

Algorithm 3 React to rebut($\neg(\alpha \to \beta)$)

```
Require: received rebut($\neg(\alpha \to \beta)$) from $A$
$\mathcal{E} \leftarrow$ set of all extensions computed from $AF$
Require: $\alpha \in \bigcup_{\mathcal{E}}, \alpha \to \beta \in \mathcal{R}$
if $B$ trusts $A$ then
    {believe what A says and revise own $AF$}
    add all attacks between $\alpha$ and $\beta$ in $A$'s $AF$
    remove all attacks between $\alpha$ and $\beta$ not in $A$'s $AF$
    compute $\mathcal{E}'$ from the revised $AF$
    if $\sigma(\mathcal{D}) \in \bigcup_{\mathcal{E}} \setminus \bigcup_{\mathcal{E}'} \cup \bigcup_{\mathcal{E}'} \setminus \bigcup_{\mathcal{E}}$ then
        {the agent changed her mind about $\sigma(\mathcal{D})$: $\mathcal{D}$ ends with agreement}
        utter ok
    else
        {deleting $\alpha \to \beta$ does not change $B$'s mind about $\sigma(\mathcal{D})$: $\mathcal{D}$ continues}
        $\mathcal{E} \leftarrow \mathcal{E}'$
        if $\exists \alpha' \to \beta \in \mathcal{R} \setminus \Delta_{out}$ such that $\alpha' \in \bigcup_{\mathcal{E}}$ then
            {there is another direct attack against $\beta$}
            utter attack($\alpha' \to \beta$)
            $\Delta_{out} \leftarrow \Delta_{out} \cup \alpha' \to \beta$
        else
            {attack something A said in the past}
            pick $\gamma \to \beta' \in \mathcal{R} \setminus \Delta_{out}$ such that $\gamma \in \bigcup_{\mathcal{E}}$ and $\beta' \to \alpha' \in \Delta_{in}$
            utter attack($\gamma \to \beta'$)
            $\Delta_{out} \leftarrow \Delta_{out} \cup \gamma \to \beta'$
        end if
    end if
else
    {do not believe A: $\mathcal{D}$ ends without an agreement}
    utter sorry
end if
```

AF, and whatever attacks between α and β are not defined in the interlocutor's *AF* will be removed from the revising agent's AF).

There is a large literature on revising beliefs in artificial intelligence and knowledge representation. In particular, work by Alchourrón et al. [1] was influential in defining a number of basic postulates (known as *AGM postulates* in the literature) that a belief revision operator should respect, in order for that operator to be considered rational. Cayrol et al. [13] propose a framework for revising an abstract *AF* along these lines. A more recent proposal for revising an abstract *AF* following a minimal change principle was put forward by Baumann [5]. However, considering the intended application of MS dialogues, which is modelling possible outcomes of human debates, respecting the AGM postulates may not be a necessary requirement after all. We plan however to investigate the application of these and other methods in MS dialogues, and evaluate which one performs best in simulating opinion diffusion in social networks.

4 Analysis

MS dialogues present interesting features from the viewpoint of behavioural modeling. They model truthful interaction, where an agent believes what she says. They embed mechanisms for epistemic vigilance, and it is possible to adjust the trade off between argumentative reasoning and trust.

All these features are relevant for our intended application of MS dialogues for social simulation (more about it in Section 5). However, it is also interesting to analyse MS dialogues from a formal viewpoint, for example, if we wish to use this type of dialogue in a context of artificial societies, populated by software agents. In this section, we briefly discuss the most significant features of MS dialogues from a more formal perspective.

Property 1. MS dialogues respect agent autonomy. In particular, if an agent A does not trust an agent B, A's participation in an MS dialogues with B will not cause any change in A's beliefs.

It is possible to see this by looking at the branches of Algorithms 1-3 that lead to an agent revising his beliefs: they are all subject to the condition **if** B trusts A.

Property 2. If a conservative belief revision operator for argumentation frameworks is used, MS dialogues do not increase the polarisation between the participants.

"Polarisation" is a notion commonly used in sociology, to measure disagreement. Intuitively, if all agents in a group think alike, the polarisation in the group is 0. If agents are split into two equally large, homogeneous groups, and the opinions of agents from the first group are maximally distant from those of the agents in the second group, polarisation in the overall group is maximum. In our framework, a possible measure of polarisation could be based on a notion of distance between AFs measured by counting the number of attacks upon which the two AFs disagree. An alternative distance could consider the overlapping in the extensions of the two AFs. The latter is the approach we follow in [20].

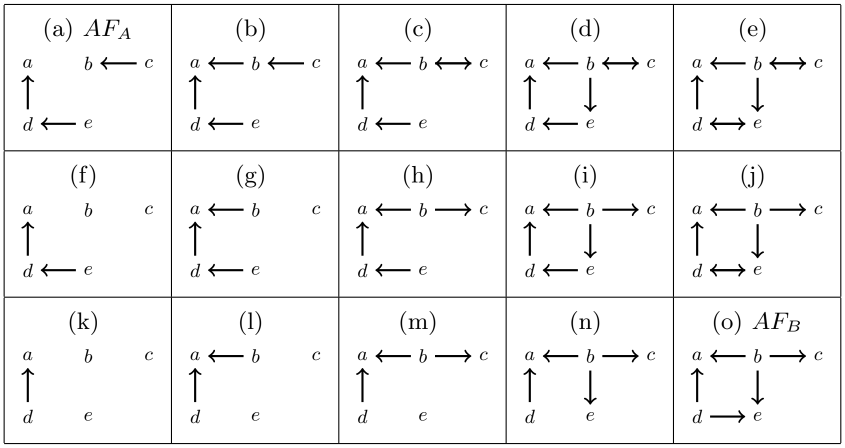
A "conservative" belief revision operator applied on a set of attack relations $\mathcal{R}_A$ during a dialogue between A and B, producing a new set of attack relations $\mathcal{R}'_A$ satisfies the following properties: (1) if $\alpha \to \beta \in \mathcal{R}_A \cap \mathcal{R}_B$, then $\alpha \to \beta \in \mathcal{R}'_A$; (2) if $\alpha \to \beta \notin \mathcal{R}_A \cup \mathcal{R}_B$, then $\alpha \to \beta \notin \mathcal{R}'_A$. The revision method we implemented, which is described at the end of Section 3, is conservative.

We give an intuitive justification of Property 2 using the example of Section 3, Fig. 1: Table 2 shows the possible states that can be reached by A or B during one or more dialogues via conservative revision steps. We can see that the maximum distance is at the initial state. This means that polarisation can only decrease.

Property 3. MS dialogues do not change focus.

The focus of an MS dialogue $\mathcal{D}$ is always its subject $\sigma(\mathcal{D})$. In particular, by Algorithm 1, either $\mathcal{D}$ terminates with agreement, or if $\mathcal{D}$ continues it does by

Table 2. Possible evolutions of AF_A and AF_B (from Section 3, Fig. 1), when interacting via MS dialogues.



attacking $\sigma(\mathcal{D})$. By Algorithm 2 and 3, all possible ramifications of the dialogue are based on attacks/denials of attacks linked to $\sigma(\mathcal{D})$, either to defeat it, or to defend it.

Property 4. The flow of dialogue is guaranteed. In particular, participants always have something to say until they decide to terminate the dialogue.

By Algorithm 1, possible utterances are *ok* (line 9) or *attack*($\alpha \rightarrow \sigma(\mathcal{D})$) (line 14). The first case is trivial. In order to execute the second option, an agent must be able to produce an attack $\alpha \rightarrow \sigma(\mathcal{D})$. Such an attack is guaranteed to exist. If it did not, $\sigma(\mathcal{D})$ would be in one of the agent's extensions, and the algorithm would not end up in this branch (line 3). Similar considerations hold for Algorithm 2 and 3.

Property 5. MS dialogues allow agents to exhaustively express all their objections to the interlocutor's claim.

Depending on their mutual (lack of) trust, there is always a possible dialogue where an agent can put forward an objection against any of the interlocutor's arguments, if there exists one that has not been put forward already.

Property 6. Given finite argumentation frameworks, MS dialogues always terminate in a finite number of steps, bounded by the squared size of the argument set.

Termination is guaranteed because the dialogue keeps track of in/out narratives ($\Delta_{in}, \Delta_{out}$), and only a finite number of attacks can be defined from a finite number of arguments. Attacks/denials of attacks cannot be repeated during the same dialogue, thus in the worst case the number of turns in a dialogue is the maximum possible size of $\mathcal{R}$, i.e., $|\{calA\}|^2$.

These properties tell us that MS dialogues can be efficiently used in simulation experiments, and that their possible outcomes follow the intuition.

5 Applying MS dialogues to social simulation

Sociologists within the Agent-Based Social Simulation (ABSS) area have attacked the mechanisms that are somewhat related to agreement, under many points of view: in terms of hierarchies, trust evolution, cooperation, opinions Polarisation and voting attitude, consensus, cultural differentiation, social structure and its effects on cooperation, cultural differentiation, norms and collective beliefs, and finally, in terms of collective behavior.

There are at least two common aspects in all these attempts: (*a*) the use of social networks to represent social embeddedness and (*b*) a preference for mathematical, game theoretical or artificial intelligence techniques.

This stream of research focuses on agents that do, in fact, interact but where very little explicit reasoning is done - and if it is, it is "compiled" into procedural code. These scholars model agent's reasoning mainly by:

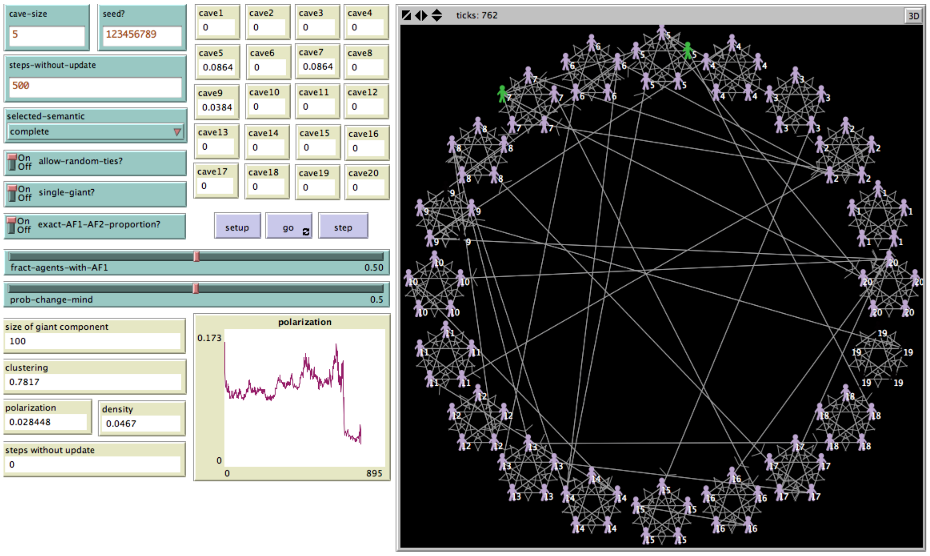
(*a*) threshold models that link the probability of an agent to choose between a set of opportunities;
(*b*) theoretical games, like the Iterated Prisoner's Dilemma, to assess the emergence of a stable regime of cooperation between bounded rational agents;
(*c*) genetic algorithms, to implement evolving collaborative or competitive strategies in game theoretical settings; or
(*d*) neural networks, to explore social meta-reasoning and beliefs.

Among the Social Sciences, this formal approach is competing with a second formal stream, which focuses explicitly on how social agents should reason *socially*, i.e., interdependently with others, by means of formal logics. The relevance of logic in ABSS is an open issue, with both detractors and supporters [16].

It is interesting to notice that BDI frameworks, like the ones advocated by Hedstrom [23], have not encountered a wide diffusion among sociologists, probably because most agent architectures based on the BDI paradigm are complex to understand and to use by non-computer-scientists, and often not suited for simulation with thousand of agents. On the other side, agents are mainly called *social* just because they are linked in network structures, but no reasoning is actually implemented.

In spite of a substantive claim for the adoption of agent-based models in the Analytical, Generative and Computational fields of Sociology, it looks like a shared framework to model key reasoning capabilities of social agents has not been developed yet (see Carley and Newell [11] for a review of possible models of social agent).

We think that argumentation technologies can be successfully used to fill this gap. Accordingly, we proposed a new model for agent-based social simulations



cave-size: 5

seed?: 123456789

steps-without-update: 500

selected-semantic: complete

On Off allow-random-ties?

On Off single-giant?

On Off exact-AF1-AF2-proportion?

cave1 0	cave2 0	cave3 0	cave4 0
cave5 0.0864	cave6 0	cave7 0.0864	cave8 0
cave9 0.0384	cave10 0	cave11 0	cave12 0
cave13 0	cave14 0	cave15 0	cave16 0
cave17 0	cave18 0	cave19 0	cave20 0

fract-agents-with-AF1: 0.50

prob-change-mind: 0.5

size of giant component: 100

clustering: 0.7817

polarization: 0.028448

density: 0.0467

steps without update: 0

Fig. 2. A snapshot of NetArg at work with a simulation. The simulation started with 50% of agents having one argumentation framework, and the other half having a different framework. When the snapshot was taken, the simulation had completed 762 steps. Notice the low Polarisation (chart on the bottom left), and the majority of agents who share the same argumentation framework (they are coloured in the same way) and thus the same opinion.

[20], whereby agents belonging to a social network reason and interact argumentatively, using MS dialogues. With this model, we simulate the propagation of arguments and evolution of opinions in a social context.

We run experiments with a simulator that implements MS dialogues [21]. The simulator, called NetArg, is developed in NetLogo [53], a standard tool in social simulation, along with a Java Extension that uses ConArg [8] to compute the semantics.

Following an experimental design due to Flache & Macy [18], we use the "disconnected caveman graph" [52] to represent a situation where components are maximally dense. The model comprises 100 agents organised in 20 "caves" (identifiable as cliques in Fig. 2): the parameter *cave-size* sets the number of agents for each cave. In order to run a simulation, two AFs must be generated, either randomly or by specifying an attack structure. Fig. 2 shows an experiment using the usual AF_A and AF_B from Section 3, Fig. 1.

Different semantics can be chosen with the *selected-semantic* parameter. AFs are thus distributed randomly with probability *fract-agents-with-AF1* among the population. Trust probability is set at a *prob-change-mind* value, which is fixed for all agents. For the sake of simplicity, we adopt a static threshold-based trust model, where the exact value of thresholds is determined by a system of linear equations aimed at guaranteeing that in some basic cases (such as the illustrations presented in this paper), the odds of a dialogue ending with A or B changing her mind are even, following a *fairness* principle. In our implementation, it is possible to specify for each simulation a “trust level”, meaning the probability that a dialogue will end with either agent changing her mind (as opposed to neither).

At each time step, each agent is asked to start a dialogue with one of its neighbors, which could be restricted to the same cave or not, depending on the presence of bridges if *allow-random-ties* is true.

NetArg can be used to answer interesting research questions for the social sciences, for example, *does the presence of bridges (or weak ties) lower the polarisation level of the population, i.e., does the model exhibit a long-range ties effect on social polarisation?*[2] We discuss some preliminary results in [20].

6 Conclusions

This article describes a new dialogue model that builds on Mercier & Sperber’s argumentative theory of reasoning and harmonises argumentation and trust. Our approach is different from others in the argumentation literature, in that it finds its motivations in the social sciences, rather than in philosophy, computer science and game theory. However, there are many related works.

Koster et a. [25] review recent literature on the combination of argumentation and trust. Most of the work done concerns using argumentation to evaluate trust, in what they call argument-supported trust [37,35,36,46,28], and arguments about trust [24]. Parsons et al. [33] investigate the combination of trust measures on agents and the use of argumentation for reasoning about belief, combining an existing system for reasoning about trust and an existing system of argumentation. This work is the most closely related to our proposal, but it does not consider dialogues. An interesting direction for future research could be towards understanding whether this framework could suit to modeling dialogues in social network, and how it relates to Sperber & Mercier’s framework.

On another line, a growing number of *debate-friendly* tools is rising to help users visualise and understand the outcome of a discussion. Among them there are (a) *visualisation* tools, such as DebateGraph;[3] (b) community-based tools relying on *user ranking*, such as DBee,[4] a global debating network which features scoring and ranking with both positive or negative values and Debate.org,[5] a social network platform where users can start a debate and comment with pro/cons rating against the main argument in the debate; community-based *moderated*, professional discussion forums. Well-known example of the last category are Debatepedia,[6] an International Debate Education Association project containing a

[2] See `http://www.youtube.com/watch?v=_YfhKpYASf0`.

[3] `http://www.debategraph.org`

[4] `http://dbeelife.com`

[5] `http://www.debate.org`

[6] `http://idebate.org/debatabase`

database of more than 500 debates, produced by professional debates and used as a training set for students who want to learn how to debate effectively and improve the database itself and Deliberatorium,[7] a community-moderated system where comments are subject to moderator approval before they may be certified and finally become visible to a larger community. Other argumentation-enabled web applications are discussed in [48,43,12,45,42]. Finally, following the bottom-up argumentation concept, *microdebates* are an argumentation-based tool devised to support online debates [19]. All these approaches have a common goal: to improve rational debates in the social Web, using argumentation technologies. Conversely, our work is primarily concerned with modeling online exchanges, and not on giving rules or tools to modify the way people communicate, and as such it is orthogonal with all the above.

Finally, we mention work done by Maudet and Bonzon on studying the outcomes of multi-party persuasion dialogues [9]. Their formal approach builds on mechanism design and is orthogonal to MS dialogues, which build on experimental psychology. That work is surely relevant to ours, and we plan to compare their theoretical results with those we present in our formal analysis, and with the experimental results we shall obtain by simulating MS dialogues using NetArg.

There are many directions for future work. In Section 2 we outline a new perspective of the meaning that can be given to abstract argumentation frameworks in social environment, where each agent/human has a different understanding of concepts that have a common abstract reference. This view tallies with the bottom-up argumentation concept [47] and with other applications of abstract argumentation in the social networks domain, such as the aforementioned microdebates [19]. Surely, more work is needed to explore this idea and understand it and how it can help bridge the gap between natural and abstract argumentation. We intend to evaluate this using data from online debates. However, as noted by Modgil et al. [32], organizing human authored arguments into Dung's argumentation frameworks is a difficult problem, on which there has been little work so far.

We have already mentioned alongside the presentation some other necessary extension to our model: towards multi-party debates, towards better-defined trust models grounded on sociological studies, and towards better-defined belief revision processes, which will have to be designed so as to best reflect the essence of Mercier & Sperber's theory. The concept of distance between AFs, used to measure polarisation, is also an important direction for further work. We are only aware of another proposal, by Booth et al. [10], to measure distance in the context of abstract argumentation, but with an entirely different perspective. Indeed, the authors propose a way to quantify disagreement, but not between two different AFs, but between two extensions inside the same AF. It would be interesting to study the problem of quantifying disagreement between AFs in general, possibly as an extension of Booth et al.'s work.

[7] `http://cci.mit.edu/research/deliberatorium.html`

Acknowledgments

We thank the anonymous reviewers for providing valuable feedback. This work was partially supported by the ePolicy EU project FP7-ICT-2011-7, grant agreement 288147. Possible inaccuracies of information are under the responsibility of the project team. The text reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained in this paper.

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet functions for contraction and revision. Journal of Symbolic Logic 50, 510–530 (1985)
2. Amgoud, L., Parsons, S., Maudet, N.: Arguments, dialogue and negotiation. In: Horn, W. (ed.) Proceedings of the Fourteenth European Conference on Artificial Intelligence, Berlin, Germany (ECAI 2000). IOS Press (Aug 2000)
3. Arieli, O.: Conflict-tolerant semantics for argumentation frameworks. In: Proc. 13th JELIA. LNCS, vol. 7519, pp. 28–40. Springer (2012)
4. Baroni, P., Giacomin, M.: Semantics of abstract argument systems. In: Argumentation in Artificial Intelligence. Springer (2009)
5. Baumann, R.: What does it take to enforce an argument? minimal change in abstract argumentation. In: Raedt, L.D., Bessière, C., Dubois, D., Doherty, P., Frasconi, P., Heintz, F., Lucas, P.J.F. (eds.) ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012. Frontiers in Artificial Intelligence and Applications, vol. 242, pp. 127–132. IOS Press (2012)
6. Bench-Capon, T.J.M.: Value-based argumentation frameworks. In: Benferhat, S., Giunchiglia, E. (eds.) 9th International Workshop on Non-Monotonic Reasoning (NMR 2002), April 19-21, Toulouse, France, Proceedings. pp. 443–454 (2002)
7. Bench-Capon, T.J.M., Dunne, P.E.: Argumentation in artificial intelligence. Artif. Intell. 171(10-15), 619–641 (2007)
8. Bistarelli, S., Santini, F.: ConArg: A constraint-based computational framework for argumentation systems. In: 23rd International Conference on Tools with Artificial Intelligence, ICTAI 2011. pp. 605–612. IEEE (2011)
9. Bonzon, E., Maudet, N.: On the outcomes of multiparty persuasion. In: Sonenberg, L., Stone, P., Tumer, K., Yolum, P. (eds.) 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), Taipei, Taiwan, May 2-6, 2011, Volume 1-3. pp. 47–54. IFAAMAS (2011)
10. Booth, R., Caminada, M., Podlaszewski, M., Rahwan, I.: Quantifying disagreement in argument-based reasoning. In: van der Hoek, W., Padgham, L., Conitzer, V., Winikoff, M. (eds.) Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012). pp. 493–500. IFAAMAS (2012)
11. Carley, K.M., Newell, A.: The nature of the social agent. J. Math. Sociol. 19(4), 221–262 (1994)
12. Cartwright, D., Atkinson, K.: Political engagement through tools for argumentation. In: Besnard, P., Doutre, S., Hunter, A. (eds.) Computational Models of Argument: Proceedings of COMMA 2008, Toulouse, France, May 28-30, 2008. Frontiers in Artificial Intelligence and Applications, vol. 172, pp. 116–127. IOS Press (2008)

13. Cayrol, C., de Saint-Cyr, F.D., Lagasquie-Schiex, M.C.: Revision of an argumentation system. In: Brewka, G., Lang, J. (eds.) Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008. pp. 124–134. AAAI Press (2008)
14. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321 – 357 (1995)
15. Dung, P.M., Mancarella, P., Toni, F.: A dialectic procedure for sceptical, assumption-based argumentation. In: Proc. 1st COMMA. Frontiers in AI and Applications, vol. 144, pp. 145–156. IOS Press (2006)
16. Edmonds, B.: How formal logic can fail to be useful for modelling or designing mas. Online resource, http://cfpm.org/logic-in-abss/papers/Edmonds.html
17. van Eemeren, F.H., Grootendorst, R.: Fallacies in pragma-dialectical perspective. Argumentation 1, 283–301 (1987)
18. Flache, A., Macy, M.W.: Small worlds and cultural polarization. J. Math. Sociol. 35(1-3), 146–176 (2011)
19. Gabbriellini, S., Torroni, P.: Large scale agreements via microdebates. In: Ossowski, S., Toni, F., Vouros, G. (eds.) Proceedings of the First International Conference on Agreement Technologies, AT 2012, Dubrovnik, Croatia, October 15-16, 2012. CEUR Conference Proceedings, vol. 918 (2012)
20. Gabbriellini, S., Torroni, P.: Arguments in social networks. In: Proceedings of the Twelfth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013). IFAAMAS (2013)
21. Gabbriellini, S., Torroni, P.: NetArg: an agent-based social simulator with argumentative agents. Under review (AAMAS 2013 DEMOs track) (2013)
22. Hamblin, C.L.: Fallacies. Methuen, London, UK (1970)
23. Hedstrom, P.: Dissecting the Social. On the Principles of Analytical Sociology. CUP (2005)
24. Koster, A., Sabater-Mir, J., Schorlemmer, M.: Personalizing communication about trust. In: Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012). pp. 655 – 664. IFAAMAS, Valencia, Spain (2012)
25. Koster, A., Sabater-Mir, J., Schorlemmer, M.: Argumentation and trust. In: Ossowski, S. (ed.) Agreement Technologies, Law, Governance and Technology Series, vol. 8, pp. 441–451. Springer Netherlands (2013), http://dx.doi.org/10.1007/978-94-007-5583-3_25
26. Macagno, F., Walton, D.: Types of dialogue, dialectical relevance, and textual congruity. Anthropology & Philosophy 8(1-2), 101–121 (2007)
27. Mackenzie, J.: Question-begging in non-cumulative systems. Journal of Philosophical Logic 8, 117–133 (1979), http://dx.doi.org/10.1007/BF00258422
28. Matt, P.A., Morge, M., Toni, F.: Combining statistics and arguments to compute trust. In: van der Hoek, W., Kaminka, G.A., Lespérance, Y., Luck, M., Sen, S. (eds.) 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), Toronto, Canada, May 10-14, 2010, Volume 1-3. pp. 209–216. IFAAMAS (2010)
29. McBurney, P., Parsons, S.: Dialogue games for agent argumentation. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 261–280. Springer US (2009), http://dx.doi.org/10.1007/978-0-387-98197-0_13
30. Mercier, H., Sperber, D.: Why do humans reason? arguments for an argumentative theory. Behavioral and Brain Sciences 34(02), 57–74 (2011), http://dx.doi.org/10.1017/S0140525X10000968

31. Modgil, S., Bench-Capon, T.J.M.: Metalevel argumentation. J. Log. Comput. 21(6), 959–1003 (2011)
32. Modgil, S., Toni, F., Bex, F., Bratko, I., Chesñevar, C.I., Dvorák, W., Falappa, M.A., Fan, X., Gaggl, S.A., García, A.J., González, M.P., Gordon, T.F., Leite, J., Molina, M., Reed, C., Simari, G.R., Szeider, S., Torroni, P., Woltran, S.: The added value of argumentation. In: Agreement Technologies, pp. 357–404. Law, Governance and Technology Series 8, Springer-Verlag (2013)
33. Parsons, S., Tang, Y., Sklar, E., McBurney, P., Cai, K.: Argumentation-based reasoning in agents with varying degrees of trust. In: Sonenberg, L., Stone, P., Tumer, K., Yolum, P. (eds.) Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), Taipei, Taiwan, May 2-6, 2011, Volume 1-3. pp. 879–886. IFAAMAS (2011)
34. Parsons, S., Wooldridge, M., Amgoud, L.: An analysis of formal inter-agent dialogues. In: The First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2002, July 15-19, 2002, Bologna, Italy, Proceedings. pp. 394–401. ACM (2002)
35. Pinyol, I., Sabater-Mir, J.: Arguing about reputation: The lrep language. In: Artikis, A., O'Hare, G.M.P., Stathis, K., Vouros, G.A. (eds.) Engineering Societies in the Agents World VIII, 8th International Workshop, ESAW 2007, Athens, Greece, October 22-24, 2007, Revised Selected Papers. Lecture Notes in Computer Science, vol. 4995, pp. 284–299. Springer (2008)
36. Pinyol, I., Sabater-Mir, J.: An argumentation-based dialog for social evaluations exchange. In: Coelho, H., Studer, R., Wooldridge, M. (eds.) ECAI 2010 - 19th European Conference on Artificial Intelligence, Lisbon, Portugal, August 16-20, 2010, Proceedings. Frontiers in Artificial Intelligence and Applications, vol. 215, pp. 997–998. IOS Press (2010)
37. Prade, H.: A qualitative bipolar argumentative view of trust. In: Prade, H., Subrahmanian, V. (eds.) Scalable Uncertainty Management, Lecture Notes in Computer Science, vol. 4772, pp. 268–276. Springer Berlin Heidelberg (2007), `http://dx.doi.org/10.1007/978-3-540-75410-7_20`
38. Prakken, H.: Logical Tools for Modelling Legal Argument. Dordrecht, Kluwer (1997)
39. Prakken, H.: Models of persuasion dialogue. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 281–300. Springer US (2009), `http://dx.doi.org/10.1007/978-0-387-98197-0_14`
40. Rahwan, I., Ramchurn, S.D., Jennings, N.R., McBurney, P., Parsons, S., Sonenberg, L.: Argumentation-based negotiation. The Knowledge Engineering Review 18(4), 343–375 (2003)
41. Sadri, F., Toni, F., Torroni, P.: Dialogues for negotiation: agent varieties and dialogue sequences. In: Intelligent Agents VIII: 8th International Workshop, ATAL 2001, Seattle, WA, USA, Revised Papers. Lecture Notes in Artificial Intelligence, vol. 2333, pp. 405–421. Springer-Verlag (2002)
42. Schneider, J., Groza, T., Passant, A.: A review of argumentation for the social semantic web. Semantic Web 4(2), 159–218 (01 2013), `http://dx.doi.org/10.3233/SW-2012-0073`
43. Shum, S.B.: Cohere: Towards web 2.0 argumentation. In: Besnard, P., Doutre, S., Hunter, A. (eds.) Computational Models of Argument: Proceedings of COMMA 2008, Toulouse, France, May 28-30, 2008. Frontiers in Artificial Intelligence and Applications, vol. 172, pp. 97–108. IOS Press (2008)
44. Skvoretz, J., Fararo, T.: Status and participation in task groups: A dynamic network model. Am. J. Sociol. 101(5), 1366–1414 (1996)

45. Snaith, M., Bex, F., Lawrence, J., Reed, C.: Implementing argublogging. In: Verheij, B., Szeider, S., Woltran, S. (eds.) Computational Models of Argument - Proceedings of COMMA 2012, Vienna, Austria, September 10-12, 2012. Frontiers in Artificial Intelligence and Applications, vol. 245, pp. 511–512. IOS Press (2012)
46. Tang, Y., Cai, K., McBurney, P., Sklar, E., Parsons, S.: Using argumentation to reason about trust and belief. J. Log. Comput. 22(5), 979–1018 (2012)
47. Toni, F., Torroni, P.: Bottom-up argumentation. In: Proc. TAFA. LNCS, vol. 7132, pp. 249–262. Springer (2012)
48. Torroni, P., Gavanelli, M., Chesani, F.: Arguing on the semantic grid. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 423–441. Springer US (2009), http://dx.doi.org/10.1007/978-0-387-98197-0_21
49. Walton, D.N., Krabbe, E.C.W.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. State University of New York Press, Albany, NY (1995)
50. Walton, D.: Types of dialogue, dialectal shifts and fallacies. In: van Eemeren, F.H., Grootendorst, R., Blair, J.A., Willard, C.A. (eds.) Argumentation Illuminated. International Society for the Study of Argumentation (SICSAT). pp. 133–147 (1992)
51. Walton, D.: Dialectical relevance in persuasion dialogue. Informal Logic 19(2), 119–143 (1999)
52. Watts, D.J.: Network dynamics and the small-world phenomenon. Am. J. Sociol. 105, 493–527 (1999)
53. Wilensky, U.: Netlogo (1999), http://ccl.northwestern.edu/netlogo/, center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.