Persuasion Dialogues via Restricted Interfaces Using Probabilistic Argumentation

Anthony Hunter(✉)

Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK
anthony.hunter@ucl.ac.uk

Abstract. For persuasion dialogues between a software system and user, a user should be able to present arguments. Unfortunately, this would involve natural language processing which is not viable for this task in the short-term. A compromise is to allow the system to present potential counterarguments to the user, and the user expresses his/her degree of belief in each of them. In this paper, we present a protocol for persuasion that supports this type of move, and show how the system can use the epistemic approach to probabilistic argumentation to model the user, and thereby optimize the choice of moves.

1 Introduction

Computational models of argument can potentially be used for systems to persuade users to change their behavior (e.g. to eat less, to exercise more, to use less electricity, to vote, to not text while driving, etc.) [14]. A **system** (the *persuader* running for example as an app) enters into a dialogue with a **user** (the *persuadee* using the app) to persuade them to believe a specific argument called the persuasion goal (e.g. eat more fruit because it is healthy for you).

By choosing appropriate arguments to present to the user, the system may raise the user's belief in the persuasion goal. However, for the system, there is a problem of how to get arguments from the user in order to support a fair and frank persuasion dialogue. We assume the system cannot understand arguments presented in natural language given the complexity of processing arguments in free text. Hence, the interface with the user is restricted. Our solution is for the system to give a menu of arguments, and the user presents agreement/disagreement in each argument by giving it a score (as in a Likert scale [20]). This score is in the unit interval and denotes the belief that the user has in the argument (i.e. the degree to which the user thinks the premises are true and the claim follows from the premises).

Example 1. Suppose the system gives argument `A` in Fig. 1 as its persuasion goal. It is aware of two potential counterarguments `B` and `C`. So it presents these in a menu, and asks the user for his/her degree of belief in them. If the user declares belief greater than 0.5 in `B` (resp. `C`), then the system presents `D` (resp. `E`) with the aim of lowering the user's belief in `B` (resp. `C`) and increasing the user's belief in `A`.

S. Schockaert and P. Senellart (Eds.): SUM 2016, LNAI 9858, pp. 184–198, 2016.
DOI: 10.1007/978-3-319-45856-4_13

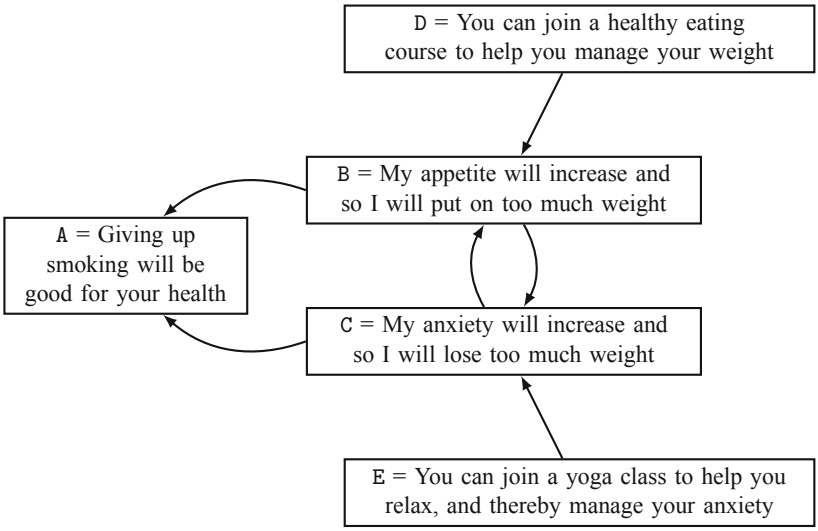


Fig. 1. Example of argument graph for persuasion. It contains the arguments known (but not necessarily believed) by the system. Argument `A` could be a persuasion goal and so `B` and `C` are potential counterarguments for the user.

The above example is a kind of asymmetric dialogue where the moves available to the persuader are different to those available to the persuadee. There is a recent proposal for asymmetric persuasion dialogues with a general definition for probabilistic user models, and a general definition for updating user models in terms of mass redistributions [16]. However, [16] does not consider the following issues: how a menu of potential counterarguments could be presented to the user, how the user could express his/her belief in each of them, or how these moves can be used in a protocol that is fair to the user. We address these issues by making the following contributions in this paper: (1) A dialogue protocol that incorporates the menu move and that is fair to the persuadee; (2) A probabilistic model of the persuadee that can be updated through the dialogue and used by the persuader to predict whether the persuasion is successful; and (3) A method for simulation of the persuadee by the persuader when deciding on which moves to make in the dialogue.

2 Dialogues via Restricted Interfaces

We base our paper on abstract argumentation [6]. We assume our dialogues concern an **argument graph** G where $\mathsf{Args}(G)$ is the set of arguments in G, and $\mathsf{Attacks}(G)$ is the set of attack relations in G. Also $\Gamma \subseteq \mathsf{Args}(G)$ is **conflict-free** iff there is no $A, B \in \Gamma$ s.t. $(A, B) \in \mathsf{Attacks}(G)$. We assume that G contains the arguments known (but not necessarily believed) by the system.

A **dialogue** is a sequence of moves $D = [m_1, \ldots, m_k]$. Equivalently, we use D as a function with an index position i to return the move at that index

(i.e. $D(i) = m_i$). A **move** is one of the following: (1) A **posit** A where $A \in \mathsf{Args}(G)$; (2) A **menu** $[A_1/X_1, \ldots, A_n/X_n]$ where for each $A/X \in [A_1/X_1, \ldots, A_n/X_n]$, $A \in \mathsf{Args}(G)$ and $X \in [0, 1]$ is the belief of the user in A; and (3) A **system termination** $\bot$.

A **protocol** specifies what moves should/can follow each move in a dialogue. For this paper, the protocol assumes that: (1) the first move is a posit called the **persuasion goal** which is the argument that the persuader wants the persuadee to believe (with a probability greater than 0.5); (2) a dialogue does not continue after the system has terminated (i.e. if $1 \leq i < k$, then $D(i) \neq \bot$); (3) each argument in a menu is a counterargument to the posit given immediately before the menu (i.e. if $D(i) = A$, and $D(i+1) = [A_1/X_1, \ldots, A_n/X_n]$, then for each $A_j/X_j \in D(i+1)$, $(A_j, A) \in \mathsf{Attacks}(G)$); and (4) the user gives the same belief to an argument if it is repeated (i.e. If $\exists i, j$ s.t. $A/X \in D(i)$ and $A/X' \in D(j)$ then $X = X'$). A dialogue D is **finite** iff $D = [m_1, \ldots, m_k]$ and k is finite.

We assume that the system controls the dialogue. At each point in the dialogue, the system makes a posit, or menu, or termination move. If it is a menu move, then the user provides his/her belief in each argument in the menu.

Example 2. For Fig. 1, if the system gives the persuasion goal $\mathtt{A}$, then $[\mathtt{B}/0.9, \mathtt{C}/0.2]$ is a menu move where $\mathtt{B}$ and $\mathtt{C}$ are from the system, and 0.9 and 0.2 are from the user.

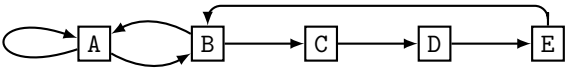
For a dialogue $D = [m_1, \ldots, m_k]$, let $\mathsf{Steps}(D) = \{1, \ldots, k\}$. For dialogues D' and D, the **subsequence relation**, denoted $D' \sqsubseteq D$, holds iff for all $i', j' \in \mathsf{Steps}(D')$, if $i' < j'$, then there are $i, j \in \mathsf{Steps}(D)$ such that $i < j$ and $D'(i') = D(i)$ and $D'(j') = D(j)$. For example, $[[\mathtt{F}/0.9, \mathtt{G}/0.2], \mathtt{D}] \sqsubseteq [\mathtt{A}, [\mathtt{F}/0.9, \mathtt{G}/0.2], \mathtt{C}, \mathtt{D}, \mathtt{E}, \bot]$. Also $D' \sqsubset D$ is defined as $D' \sqsubseteq D$ and not $D \sqsubseteq D'$.

3 Fair Dialogues

In this section, we ensure dialogues are fair by allowing the persuadee to express belief in potential counterarguments.

Definition 1. *For $A, B \in \mathsf{Args}(G)$, A **indirectly attacks** B iff (1) $A \neq B$ and (2) either $(A, B) \in \mathsf{Attacks}(G)$ or there are $(A, A'), (A', A'') \in \mathsf{Attacks}(G)$ s.t. $A \neq A'$ and $A' \neq A''$ and A'' indirectly attacks B.*

Example 3. Let $\rightsquigarrow$ denote the "indirectly attacks" relationship. So for the following graph $\mathtt{A} \rightsquigarrow \mathtt{B}$, $\mathtt{A} \rightsquigarrow \mathtt{D}$, $\mathtt{B} \rightsquigarrow \mathtt{A}$, $\mathtt{B} \rightsquigarrow \mathtt{C}$, $\mathtt{B} \rightsquigarrow \mathtt{E}$, $\mathtt{C} \rightsquigarrow \mathtt{D}$, $\mathtt{C} \rightsquigarrow \mathtt{B}$, $\mathtt{D} \rightsquigarrow \mathtt{A}$, $\mathtt{D} \rightsquigarrow \mathtt{E}$, $\mathtt{D} \rightsquigarrow \mathtt{C}$, $\mathtt{E} \rightsquigarrow \mathtt{B}$, and $\mathtt{E} \rightsquigarrow \mathtt{D}$.



Proposition 1. *Let $X \subseteq \mathsf{Args}(G)$ be s.t. there is no $A \in X$ where $(A, A) \in \mathsf{Attacks}(G)$. X is conflict-free iff for all $A, B \in X$, it is not the case that A indirectly attacks B.*

Definition 2. *For $A, B \in \mathsf{Args}(G)$, A* **defends** *B iff (1) $A \neq B$ and (2) either there is a $(A, C), (C, B) \in \mathsf{Attacks}(G)$ s.t. $A \neq C$ and $C \neq B$, or there is a C s.t. A defends C and C defends B.*

Proposition 2. *For $B, A \in \mathsf{Args}(G)$, if B indirectly attacks A, then there is a $(B, C) \in \mathsf{Attacks}(G)$ s.t. $C = A$ or C defends A.*

To compose the menus, we assume in Definition 4 that each posit is followed by a menu of arguments that attack the posit according to the argument graph, and that have not already appeared in a menu and indirectly attacked by the posit. As we cover in Sect. 5, we will aim for belief in the posit and disbelief in the counterargument, and so informally, if a posit indirectly attacks a counterargument in an earlier menu, then we do not need to present it to the user in a menu again.

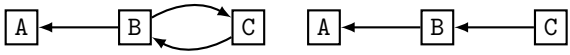
Definition 3. *For a dialogue D, a graph G, an argument A, and a step i. The* **fair attacks**, $\mathsf{FairAttacks}(G, D, A, i)$, *is $\{B \mid (B, A) \in \mathsf{Attacks}(G)$ and there is no $j < i$ s.t. $B/Y \in D(j)$ and A indirectly attacks $B\}$.*

Definition 4. *A dialogue D is* **fair** *for G iff for each i,*

$$\begin{array}{l} \textit{if } D(i) = A \textit{ and } \mathsf{FairAttacks}(G, D, A, i) \neq \emptyset \\ \textit{then } D(i+1) = [B_1/X_1, \ldots, B_n/X_n] \end{array}$$

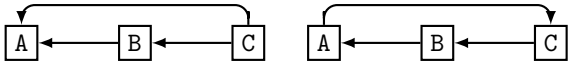
where $\mathsf{FairAttacks}(G, D, A, i) = \{B_1, \ldots, B_n\}$.

Example 4. The dialogue $[\mathtt{A}, [\mathtt{B}/0.9], \mathtt{C}, \bot]$ is fair for both the following graphs.

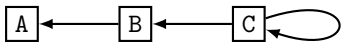


Example 5. For Fig. 1, $[\mathtt{A}, [\mathtt{B}/1, \mathtt{C}/0], \mathtt{D}, \bot]$, $[\mathtt{A}, [\mathtt{B}/0, \mathtt{C}/0.7], \mathtt{E}, \bot]$, $[\mathtt{A}, [\mathtt{B}/0, \mathtt{C}/0], \bot]$, $[\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/1], \mathtt{C}, [\mathtt{B}/0.9], \bot]$, and $[\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/0.65], \mathtt{D}, \mathtt{E}, \bot]$, are fair.

Example 6. The dialogue $[\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/0.7], \mathtt{C}, \bot]$ is fair for the left graph and the dialogues $[\mathtt{A}, [\mathtt{B}/0.5], \bot]$ and $[\mathtt{A}, [\mathtt{B}/1], \mathtt{C}, [\mathtt{A}/0.9], \mathtt{B}, [\mathtt{C}/0.9], \mathtt{A}, [\mathtt{B}/1], \ldots]$ are fair for the right graph.



Example 7. For the following graph, C does not indirectly attack C and so the self-attacks causes the fair dialogue $[\mathtt{A}, [\mathtt{B}/1, \mathtt{C}/1], \mathtt{C}, [\mathtt{C}/1], \mathtt{C}, [\mathtt{C}/1], \ldots]$ to be infinite.



An **odd cycle** is a sequence of arguments $A_1, \ldots, A_m$ s.t. each A_{i+1} attacks A_i and A_1 attacks A_m where m is odd.

Proposition 3. *If argument graph G contains no odd cycles, and D is a fair dialogue, then D is finite.*

We can assign responsibility of arguments to the persuadee and persuader as follows.

Definition 5. *Let D be a dialogue, the* **persuader arguments** *are* $\mathsf{Persuader}(D) = \{A \mid \exists i \in \mathsf{Steps}(D)$ *s.t.* $D(i) = A\}$ *and the* **persuadee arguments**, *are* $\mathsf{Persuadee}(D) = \{B \mid \exists i \in \mathsf{Steps}(D)$ *s.t.* $B/X \in D(i)\}$.

Example 8. For $D = [\mathtt{A}, [\mathtt{B}/0.9], \mathtt{C}, \bot]$, $\mathsf{Persuader}(D) = \{\mathtt{A}, \mathtt{C}\}$ and $\mathsf{Persuadee}(D) = \{\mathtt{B}\}$.

From the perspective of the user, if the dialogue is fair, then s/he has been able to express his/her belief/disbelief in the potential counterarguments known by the system.

4 Probabilistic User Models

We use the epistemic approach to probabilistic argumentation [1,13,17,25].

Definition 6. *A* **mass distribution** P *over* $\mathsf{Args}(G)$ *is such that* $\sum_{\Gamma \subseteq \mathsf{Args}(G)} P(\Gamma) = 1$. *Let* $\mathsf{Dist}(G)$ *be the set of mass distributions over G. The* **probability of an argument** A *is* $P(A) = \sum_{\Gamma \subseteq \mathsf{Args}(G)}$ *s.t.* $_{A \in \Gamma} P(\Gamma)$.

For a mass distribution P, and $A \in \mathsf{Args}(G)$, $P(A)$ is the belief that an agent has in A (i.e. the degree to which the agent believes the premises and the conclusion drawn from those premises). When $P(A) > 0.5$, then the agent believes the argument to some degree, whereas when $P(A) \leq 0.5$, then the agent disbelieves the argument to some degree.

Definition 7. *The* **epistemic extension** *for mass distribution P is* $\mathsf{Extension}(P) = \{A \in \mathsf{Args}(G) \mid P(A) > 0.5\}$.

Example 9. Consider the graph in Fig. 1. If $P(\mathtt{A}) = 0.2$, $P(\mathtt{B}) = 0.9$, $P(\mathtt{C}) = 0.4$, $P(\mathtt{D}) = 0.2$, and $P(\mathtt{E}) = 0.8$, then $\mathsf{Extension}(P) = \{\mathtt{B}, \mathtt{E}\}$.

The epistemic approach provides a finer grained assessment of an argument graph than given by Dung's definition of extensions. By adopting constraints on the distribution, the epistemic approach subsumes Dung's approach [25]. However, there is also a need for a non-standard view [17] where we adopt weaker constraints on the distribution. For instance, an important aspect of the epistemic approach is the representation of disbelief in arguments even when they are unattacked. In this case, a key constraint for the non-standard view is the following which ensures that the mass distribution respects the structure of the graph, without forcing an unattacked argument to be believed [13].

Definition 8. *A mass distribution P is* **rational** *for G iff* $\forall (A, B) \in \mathsf{Attacks}(G)$, *if* $P(A) > 0.5$, *then* $P(B) \leq 0.5$.

	A	B	C	D	E	Rational
P_1	0.6	0.9	0.4	0.6	0.7	No
P_2	0.3	0.9	0.3	0.1	0.8	Yes
P_3	0.9	0.1	0.2	0.8	0.2	Yes

Example 10. Examples of mass distribution for Fig. 1.

The system (the persuader) uses a mass distribution as a model of the user (the persuadee). It can update the model at each stage of the dialogue. This is useful for asymmetric dialogues where the user is not allowed to posit arguments/counterarguments. So the only way the user can treat arguments that s/he does not accept is by disbelieving them (and the user model is intended to reflect this). In contrast, in symmetric dialogues, the user can posit counterarguments to an argument that s/he does not accept.

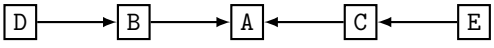
5 Winning Dialogues

In this paper, we consider two mass distributions for a dialogue. The first is the **initial distribution**, denoted P_0, which is the model of the user before the dialogue starts, and the second is the **final distribution**, denoted P_k which is the model of the user once the dialogue of k steps has terminated. In this section, we assume we have the final distribution, and in Sect. 7 we discuss how the final distribution can be obtained from the initial distribution using the moves in the dialogue.

The next definition ensures that every menu item that is changed from believed (when the user presents belief in the menu item) to disbelieved (by the end of the dialogue) has an attacker that is posited later in the dialogue and is believed.

Definition 9. *A dialogue D is* **frank** *for final distribution P_k iff for $1 \leq i \leq k$, for each $B/X \in D(i)$, if $X > 0.5$, and $P_k(B) \leq 0.5$, then there is an index j and argument C such that $i < j$ and $D(j) = C$ and $(C, B) \in \mathsf{Attacks}(G)$ and $P_k(C) > 0.5$ and $C \neq B$.*

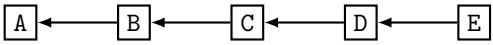
Example 11. The dialogue $[\mathtt{A}, [\mathtt{B}/1, \mathtt{C}/0.8], \mathtt{D}, \mathtt{E}, \bot]$ is fair and frank for the following argument graph G and rational final distribution P_k where $P_k(\mathtt{A}) = 0.8$, $P_k(\mathtt{B}) = 0.2$, $P_k(\mathtt{C}) = 0.2$, $P_k(\mathtt{D}) = 0.9$, and $P_k(\mathtt{E}) = 0.9$.



From the perspective of the persuader, if s/he wants to persuade the persuadee of the persuasion goal A, then the aim is for $P_k(A) > 0.5$ where P_k is the final distribution, and so the persuader can regard the dialogue as a winning dialogue, whereas if $P_k(A) \leq 0.5$, then the persuader can regard the dialogue as a losing dialogue. We formalize this next.

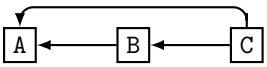
Definition 10. *Let P_k be a rational final distribution, and let D be a fair, finite, and frank, dialogue w.r.t. P_k and G s.t. $D(1) = A$ and $D(k) = \bot$. If $P_k(A) > 0.5$, then D is a* **winning dialogue***, otherwise D is a* **losing dialogue***.*

Example 12. For the following argument graph G and rational mass distribution P_k where $P_k(\mathtt{A}) = 0.9$, $P_k(\mathtt{B}) = 0$, $P_k(\mathtt{C}) = 1$, $P_k(\mathtt{D}) = 0$, and $P_k(\mathtt{E}) = 0.6$.



Let $D = [\mathtt{A}, [\mathtt{B}/0.9], \mathtt{C}, [\mathtt{D}/0.6], \mathtt{E}, \bot]$. So D is fair, finite, and frank for P_k, and D is a winning dialogue. Also $\mathsf{Persuader}(D) = \{\mathtt{A}, \mathtt{C}, \mathtt{E}\}$ and $\mathsf{Persuadee}(D) = \{\mathtt{B}, \mathtt{D}\}$.

Example 13. For the following argument graph G and rational final distribution P_k where $P_k(\mathtt{A}) = 0$, $P_k(\mathtt{B}) = 0$, and $P_k(\mathtt{C}) = 1$.



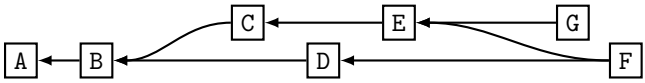
Let $D = [\mathtt{A}, [\mathtt{B}/0, \mathtt{C}/1], \mathtt{C}, \bot]$. So D is fair, finite, and frank for P_k, and D is a losing dialogue. Also $\mathsf{Persuader}(D) = \{\mathtt{A}, \mathtt{C}\}$ and $\mathsf{Persuadee}(D) = \{\mathtt{B}, \mathtt{C}\}$.

Example 14. For the graph in Fig. 1 and rational distribution P_k where $P_k(\mathtt{A}) = 0.7$, $P_k(\mathtt{B}) = 0$, $P_k(\mathtt{C}) = 0$, $P_k(\mathtt{D}) = 1$, and $P_k(\mathtt{E}) = 1$. Let $D = [\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/0.8], \mathtt{D}, \mathtt{E}, \bot]$. So D is fair, finite, and frank for P_k, and D is a winning dialogue. Also $\mathsf{Persuader}(D) = \{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ and $\mathsf{Persuadee}(D) = \{\mathtt{B}, \mathtt{C}\}$.

We now introduce the notion of minimality of a dialogue to remove superfluous moves.

Definition 11. *Let D be a winning dialogue w.r.t. P_k and G. D is* **minimal** *iff for all $D' \sqsubseteq D$, D' is not a winning dialogue w.r.t. P_k and G.*

Example 15. Fair dialogues for the graph include $D_1 = [\mathtt{A}, [\mathtt{B}/0.8], \mathtt{C}, [\mathtt{E}/0.9], \mathtt{F}, \bot]$, $D_2 = [\mathtt{A}, [\mathtt{B}/0.8], \mathtt{D}, [\mathtt{F}/0.9], \bot]$, and $D_3 = [\mathtt{A}, [\mathtt{B}/0.8], \mathtt{C}, [\mathtt{E}/0.9], \mathtt{F}, \mathtt{G}, \bot]$. Let $P_k(\mathtt{A}) = 0.8$, $P_k(\mathtt{B}) = 0$, $P_k(\mathtt{C}) = 0.8$, $P_k(\mathtt{D}) = 0$, $P_k(\mathtt{E}) = 0$, $P_k(\mathtt{F}) = 0.8$, and $P_k(\mathtt{G}) = 0.8$. So D_1 and D_3 are winning. D_2 is not frank and so losing. Also D_1 is minimal but D_3 is not minimal.



The following results show that minimal winning dialogues are well-behaved in that (1) the persuader arguments are conflict-free, (2) each persuadee argument is either not believed by the persuadee (as indicated in the menu) or is countered by the persuader, (3) the persuader and persuadee arguments are disjoint, and (4) all persuader arguments are believed and no persuadee argument is believed.

Proposition 4. *Let G be an argument graph and P_k be a rational final distribution. If D is a minimal winning dialogue w.r.t. P_k and G, then* $\mathsf{Persuader}(D)$ *is conflict-free.*

Proposition 5. *Let G be an argument graph and P_k be a rational final distribution. Also let D be a minimal winning dialogue w.r.t. P_k and G. For all $(B, A) \in \mathsf{Attacks}(G)$, if $A \in \mathsf{Persuader}(D)$, then either $B/X \in D(i)$ for some i and $X \leq 0.5$ or there is $C \in \mathsf{Persuader}(D)$ s.t. $(C, B) \in \mathsf{Attacks}(G)$.*

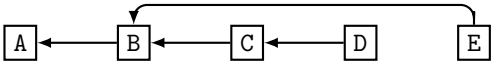
Note, we do not assume that the user is always consistent. For example, in Fig. 1, the final distribution could be s.t. $P_k(\mathtt{B}) = 0.9$ and $P_k(\mathtt{C}) = 0.8$. This would give $\mathsf{Extension}(P_k) = \{\mathtt{B}, \mathtt{C}\}$ which is not conflict-free. Of course, this would mean that the dialogue is not a winning dialogue for the persuader.

Proposition 6. *Let G be an argument graph and P be a rational final distribution. If D is a minimal winning dialogue w.r.t. P_k and G, then* $\mathsf{Persuader}(D) \cap \mathsf{Persuadee}(D) = \emptyset$.

Proposition 7. *Let G be an argument graph and P_k be a rational final distribution. If D is a minimal winning dialogue w.r.t. P_k and G, then for all $A \in \mathsf{Persuader}(D)$, $P_k(A) > 0.5$ and for all $B \in \mathsf{Persuadee}(D)$, $P_k(B) \leq 0.5$.*

The following example shows that a winning dialogue does not necessarily have all its persuader arguments being in the epistemic extension.

Example 16. Consider the following graph with final distribution $P_k(\mathtt{A}) = 1$, $P_k(\mathtt{B}) = 0$, $P_k(\mathtt{C}) = 0$, $P_k(\mathtt{D}) = 1$, and $P_k(\mathtt{E}) = 1$. So $\mathsf{Extension}(P_k) = \{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$. The dialogue $D = [\mathtt{A}, [\mathtt{B}/1], \mathtt{C}, [\mathtt{D}/1], \mathtt{E}, \bot]$ is winning w.r.t. P_k and G. Also $\mathsf{Persuader}(D) = \{\mathtt{A}, \mathtt{C}, \mathtt{E}\}$. So the persuader arguments are not a subset of the epistemic extension. However, $D' = [\mathtt{A}, [\mathtt{B}/1], \mathtt{E}, \bot]$ is a subdialogue where $\mathsf{Persuader}(D') \subseteq \mathsf{Extension}(P_k)$ and it is winning w.r.t. P_k and G.



Proposition 8. *If P_k is a rational final distribution, and D is a minimal winning dialogue w.r.t. P_k and G, then* $\mathsf{Persuader}(D) \subseteq \mathsf{Extension}(P_k)$ *holds.*

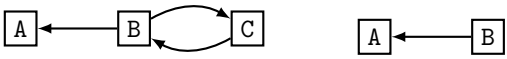
So a minimal dialogue uses arguments in the epistemic extension of P_k to present a winning position for the goal.

6 Delineated Subgraphs

The aim of this section is to better understand the proposal so far. For this, we consider properties of the subgraph of the argument graph as delineated by the dialogue.

Definition 12. *Let D be a dialogue and let G' be a subgraph of G. D* **delineates** *G' iff* $\mathsf{Args}(G') = \{A \mid \exists i \text{ s.t. } D(i) = A \text{ or } A/X \in D(i)\}$ *and* $\mathsf{Attacks}(G') = \{(A, B) \in \mathsf{Attacks}(G) \mid A, B \in \mathsf{Args}(G')\}$.

Example 17. For the following graph (left), $D_1 = [\mathtt{A}, [\mathtt{B}/1], \mathtt{C}, [\mathtt{B}/1], \mathtt{C}, \ldots]$ delineates the graph (left), whereas $D_2 = [\mathtt{A}, [\mathtt{B}/1], \bot]$ delineates the subgraph (right).



So when a dialogue D delineates a graph G, the nodes in G are exactly the arguments that appear in the posits and menus of D, and the arcs are just the arcs from the argument graph that involve those arguments.

A user declaration is what a user initially believes in an argument in a menu. Only some arguments have a user declaration, and the aim of the dialogue is to change the user's beliefs in some of these user declarations in order to have a winning dialogue.

Definition 13. *For a dialogue D, let* $\mathsf{Declarations}(D) = \{B/X \mid \exists i \text{ s.t. } B/X \in D(i)\}$ *be the arguments in a menu, let* $\mathsf{Declared}(D) = \{B \mid B/X \in \mathsf{Declarations}(D)\}$ *and let* $\mathsf{Undeclared}(D) = \{A \in \mathsf{Args}(G) \mid A \notin \mathsf{Declared}(D)\}$.

Example 18. Consider the graph in Fig. 1. For the dialogue $[\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/0.1], \mathtt{D}, \bot]$, we get $\mathsf{Declared}(D) = \{\mathtt{B}, \mathtt{C}\}$ and $\mathsf{Undeclared}(D) = \{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$.

The next definition retrieves the belief that the user assigns to each argument in a menu, and assigns belief of 0 to any argument that does not appear in a menu.

Definition 14. *The* **declared belief***, denoted Q_D, of the persuadee in dialogue D is*

$$Q_D(B) = \begin{cases} X & \text{for each } B/X \in \mathsf{Declarations}(D) \\ 0 & \text{for each } B \in \mathsf{Undeclared}(D) \end{cases}$$

Example 19. Continuing Example 18, $Q_D(\mathtt{A}) = 0$, $Q_D(\mathtt{B}) = 0.9$, $Q_D(\mathtt{C}) = 0.1$, $Q_D(\mathtt{D}) = 0$, and $Q_D(\mathtt{E}) = 0$.

The following definition captures the subgraph of argument graph G that contains all the relevant arguments given the user beliefs. It is based on a partition of the nodes in the subgraph. One partition denotes the persuader arguments and the other partition denotes the persuadee arguments. Essentially, for each persuader argument in the subgraph, all the attackers of the argument are also in the subgraph, whereas for each persuadee argument in the subgraph, all the attackers of the argument are also in the subgraph, or the persuadee argument is not believed by the persuadee.

Definition 15. *Let Q_D be the declared belief in D. $G' \sqsubseteq G$ is a* **good subgraph** *of G for D iff there is a partition of* $\mathsf{Args}(G')$ *into sets Φ and Ψ (i.e. $\Phi \cap \Psi = \emptyset$ and $\Phi \cup \Psi = \mathsf{Args}(G')$), such that the persuasion goal is in Φ, and for each $A \in \Phi \cup \Psi$,*

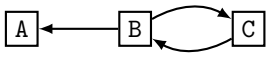
- *if* $A \in \Psi$, *then* $Q_D(A) \leq 0.5$ *or* $\exists (B, A) \in \mathsf{Attacks}(G)$ *s.t.* $(B \in \Phi$ *and* $(B, A) \in \mathsf{Attacks}(G'))$
- *if* $A \in \Phi$, *then* $\forall (B, A) \in \mathsf{Attacks}(G)$, $(B \in \Psi$ *and* $(B, A) \in \mathsf{Attacks}(G'))$

We call (Φ, Ψ) *the* **partition** *of the good subgraph.*

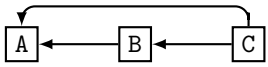
So a good subgraph is identified just by the declared beliefs expressed by the user in the menu moves. As shown below, not every fair dialogue has a good subgraph.

Example 20. The dialogue $[\mathtt{A}, [\mathtt{B}/1, \mathtt{C}/1], \mathtt{D}, \mathtt{E}, \bot]$ is winning for Fig. 1 and the final distribution P_k where $P_k(\mathtt{A}) = 1$, $P_k(\mathtt{B}) = 0$, $P_k(\mathtt{C}) = 0$, $P_k(\mathtt{D}) = 1$, and $P_k(\mathtt{E}) = 1$. The graph is the good subgraph for D with partition $\Phi = \{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ and $\Psi = \{\mathtt{B}, \mathtt{C}\}$.

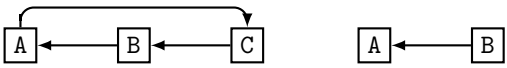
Example 21. The dialogue $[\mathtt{A}, [\mathtt{B}/1], \mathtt{C}, \bot]$ is winning for the following graph and the final distribution P_k where $P_k(\mathtt{A}) = 1$, $P_k(\mathtt{B}) = 0$, and $P_k(\mathtt{C}) = 1$. The graph is the good subgraph for D with partition $\Phi = \{\mathtt{A}, \mathtt{C}\}$ and $\Psi = \{\mathtt{B}\}$.



Example 22. Dialogues $[\mathtt{A}, [\mathtt{B}/1, \mathtt{C}/1], \mathtt{C}, \bot]$ and $[\mathtt{A}, [\mathtt{B}/1, \mathtt{C}/0], \mathtt{C}, \bot]$ are losing for the graph and any final rational distribution. There is no good subgraph for the above dialogues, whereas the dialogue $[\mathtt{A}, [\mathtt{B}/0.3, \mathtt{C}/0.1], \bot]$ is winning for the graph and a good subgraph (which is the graph itself) has the partition $\Phi = \{\mathtt{A}\}$ and $\Psi = \{\mathtt{B}, \mathtt{C}\}$.



Example 23. $[\mathtt{A}, [\mathtt{B}/1], \mathtt{C}, [\mathtt{A}/1], \mathtt{B}, [\mathtt{C}/1], \mathtt{A}, [\mathtt{B}/1], \ldots]$ is a losing dialogue for the graph (left), and any rational final distribution. There is no good subgraph for the above dialogue, whereas the dialogue $[\mathtt{A}, [\mathtt{B}/0], \bot]$ is winning for the graph and its good subgraph (right) has the partition $\Phi = \{\mathtt{A}\}$ and $\Psi = \{\mathtt{B}\}$.



Next we show that the partition of a good subgraph splits the arguments between persuader and persuadee.

Proposition 9. *If* D *is a winning dialogue w.r.t.* P_k *and* G *and* (Φ, Ψ) *is the partition of the good subgraph of* G *for* D, *then* $\Phi = \mathsf{Persuader}(D)$ *and* $\Psi = \mathsf{Persuadee}(D)$.

The following result shows that if the persuasion goal of dialogue D is believed (according to the final distribution P_k), and G' is a good subgraph of G for D, then G' does not contain any odd cycles.

Proposition 10. *If G' is a good subgraph of G for D, then G' contains no odd cycles.*

We now consider how the declarative notion of a good subgraph corresponds to winning dialogues (and the associated delineated subgraph). We show that we get a good subgraph from a minimal winning dialogue, and then we show that we can construct a winning dialogue from a good subgraph.

Proposition 11. *Let $D(1) = A$. If D is a minimal winning dialogue w.r.t. P_k and G, then there is a G' s.t. G' is a good subgraph of G for D where D delineates G' and P_k is rational for G' and $P_k(A) > 0.5$.*

Proposition 12. *If G' is a good subgraph of G for D, where (Φ, Ψ) is the partition of G', and P_k is a mass distribution s.t. $P_k(B) > 0.5$ for each $B \in \Phi$, and $P_k(C) \leq 0.5$ for each $C \in \Psi$, then there is a dialogue D, where D is a winning dialogue w.r.t. P_k and G, and D delineates G'.*

So the notion of the good subgraph provides a declarative perspective on winning dialogues.

7 Updating Mass

Given an initial distribution P_0, representing the system's model of the user's beliefs at the start of the dialogue, we update the model to give the final distribution P_k. For this, we introduce the notion of an update method which generates a mass distribution P_k from P_0 based on the moves in D.

Definition 16. *Let P_0 be an initial distribution and let D be a dialogue. An* **update function**, $\mathsf{Update}(P_0, D)$, *returns a final distribution P_k such that if $D = [\bot]$, then $P_0 = P_k$.*

There are many possibilities for defining an update function. Here we give a basic update function (below) as an example. It updates the belief in an argument based on the belief in the arguments appearing after it in the dialogue. For $D(i) = A$, belief in the arguments in the menu $D(i+1) = [B_1/X_1, \ldots, B_n/X_n]$ determines the belief in A. Similarly, for $D(i) = [B_1/X_1, \ldots, B_n/X_n]$, and each B_j in the menu, belief in the posits that occur after move $D(i)$ (i.e. moves that occur from $i+1$ to k) determine the belief in B_j.

Definition 17. *For initial distribution P_0 and dialogue D, a* **basic update function** *is $\mathsf{Update}(P_0, D) = P_k$ s.t. for each $A \in \{B \mid \exists i \text{ s.t. } D(i) = B \text{ or } B/X \in D(i)\}$:*

$$P_k(A) = \begin{cases} 0.2 & \text{if } A \in \mathsf{Persuader}(D) \text{ and } \exists B \in \mathsf{Opp}(D, A) \text{ s.t. } P_k(B) > 0.5 \\ 0.2 & \text{if } A \in \mathsf{Persuadee}(D) \text{ and } \exists B \in \mathsf{Pro}(D, A) \text{ s.t. } P_k(B) > 0.5 \\ 0.8 & \text{if } A \in \mathsf{Persuader}(D) \text{ and } \forall B \in \mathsf{Opp}(D, A), P_k(B) \leq 0.5 \\ Q_D(A) & \text{if } A \in \mathsf{Persuadee}(D) \text{ and } \forall B \in \mathsf{Pro}(D, A), P_k(B) \leq 0.5 \end{cases}$$

where $\mathsf{Opp}(D, A) = \{B \mid \exists i \text{ s.t. } D(i) = A \text{ and } B/X \in D(i+1)\}$ *and* $\mathsf{Pro}(D, A) = \{B \mid \exists i, j \text{ s.t. } i < j \text{ and } A/X \in D(i) \text{ and } D(j) = B \text{ and } (B, A) \in \mathsf{Attacks}(G)\}$.

Example 24. Consider the graph in Fig. 1. For $D = [\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/0.4], \mathtt{D}, \bot]$, with $P_0(\mathtt{A}) = 0.1$, $P_0(\mathtt{B}) = 0.7$, $P_0(\mathtt{C}) = 0.5$, $P_0(\mathtt{D}) = 0.1$, and $P_0(\mathtt{E}) = 0.1$. For the basic update function, $\mathsf{Update}(P_0, D) = P_k$ where $P_k(\mathtt{A}) = 0.8$, $P_k(\mathtt{B}) = 0.2$, $P_k(\mathtt{C}) = 0.4$, $P_k(\mathtt{D}) = 0.8$, and $P_k(\mathtt{E}) = 0.1$.

The values 0.2 and 0.8 in the basic update definition are indicative of possible assignments. More sophisticated modelling of users allows for the calculation of the value as a function of the value assigned to the counterarguments.

Proposition 13. *If* $\mathsf{Update}(P_0, D) = P_k$ *is basic, and* D *delineates* G'*, then* P_k *is rational for* G'*.*

There is a range of alternatives to the basic update in [16] that allow for a range of different kinds of user to be modelled. These include options for modelling more credulous and more skeptical users.

8 Using a User Model to Optimize Dialogues

The system wants a final distribution P_k s.t. $P_k(A) > 0.5$ for persuasion goal A. This is done in one of two modes.

In **interaction mode**, the system gives posit and menu moves, and the user gives belief in each argument in each menu (as in Example 24). At the end of the dialogue, the final mass P_k is obtained using an update function, and $P_k(A)$ is used as a prediction of the degree to which the user believes the persuasion goal $D(1) = A$.

In **simulation mode**, the system simulates a dialogue with the user in order to predict the outcome. For this, the initial mass P_0 is used for the user responses (and so P_0 is a proxy for the user answers). If this simulation is run with each possible dialogue, a dialogue can be chosen that maximizes $P_k(A)$ where A is the persuasion goal.

In this section, we focus on simulation mode. For optimization, we consider the fair and finite dialogues for a particular persuasion goal A and initial mass P_0. We denote this set $\mathsf{Fair}(G, A, P_0)$. The set of simulated dialogues is the subset where each user response is specified by the initial distribution. We use the simulated dialogues when we consider what would be the optimal choice of dialogue before undertaking the actual dialogue.

Definition 18. *The set of* **simulated dialogues**, *denoted* $\mathsf{Simulate}(G, A, P_0)$, *is* $\{D \in \mathsf{Fair}(G, A, P_0) \mid$ *for each* i, *if* $B/X \in D(i)$, *then* $P_0(B) = X\}$.

Example 25. Consider Fig. 1 with the initial distribution P_0 where $P_0(\mathtt{A}) = 0.2$, $P_0(\mathtt{B}) = 0.9$, $P_0(\mathtt{C}) = 0.7$, $P_0(\mathtt{D}) = 0.1$, and $P_0(\mathtt{E}) = 0.5$. So the fair dialogue $[\mathtt{A}, [\mathtt{B}/0.9, \mathtt{C}/0.7], \mathtt{D}, \mathtt{E}, \bot]$ is a simulated dialogue.

Definition 19. *For a dialogue* D*, with the initial distribution* P_0*, a basic update function* $\mathsf{Update}(P_0, D) = P_k$*, and persuasion goal* $D(i) = A$*, the* **score function** *is defined as* $\mathsf{Score}(D, P_0) = P_k(A)$.

Example 26. For a basic update function with Example 25, $\mathsf{Score}(D, P_0) = 0.8$.

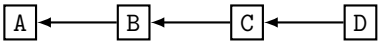
We define the locally optimal dialogues as dialogues for which all subdialogues have a lower score, and all superdialogues do not have a higher score. So a locally optimal dialogue is minimal in the sense that every move in the dialogue is required in order to get the score, and it is minimal in the sense that adding further moves will not improve the score.

Definition 20. *The* **locally optimal dialogues** *are the dialogues* $\mathsf{Local}(G, A, P_0) = \{D \in \mathsf{Simulate}(G, A, P_0) \mid \forall D' \in \mathsf{Simulate}(G, A, P_0)$, *if* $D' \sqsubset D$, *then* $\mathsf{Score}(D', P_0) < \mathsf{Score}(D, P_0)$ *and if* $D \sqsubset D'$, *then* $\mathsf{Score}(D', P_0) \leq \mathsf{Score}(D, P_0)\}$.

A globally optimal dialogue is a locally optimal dialogue that has the maximum score of locally optimal dialogues.

Definition 21. *The* **globally optimal dialogues** *are the dialogues* $\mathsf{Global}(G, A, P_0) = \{D \in \mathsf{Local}(G, A, P_0) \mid \forall D' \in \mathsf{Local}(G, A, P_0) \quad \mathsf{Score}(D', P_0) \leq \mathsf{Score}(D, P_0)\}$.

Example 27. For the following graph, let $P_0(\mathtt{A}) = 0.6$, $P_0(\mathtt{B}) = 0.3$, $P_0(\mathtt{C}) = 0.3$, and $P_0(\mathtt{D}) = 0.9$.



The final distribution P_k for each dialogue is given below. So D_1 and D_2 are winning dialogues for P_k, but only D_2 is locally optimal (and therefore globally optimal).

	A	B	C	D
$D_1 = [\bot]$	0.6	0.3	0.3	0.9
$D_2 = [\mathtt{A}, [\mathtt{B}/0.3], \bot]$	0.8	0.3	0.3	0.9

Proposition 14. *If there is a winning dialogue* D *for* G *and* P_k, *where* $\mathsf{Update}(P_0, D) = P_k$, *then there is a* $D' \in \mathsf{Global}(G, A, P_0)$ *s.t.* $\mathsf{Score}(D', P_0) > 0.5$.

So if there is a winning dialogue, then there is a globally optimal dialogue with the same outcome.

9 Discussion

In this paper, we have made the following contributions: (1) Introduced the menu move to get the user's belief in potential counterarguments; (2) Presented a fair and frank protocol for persuasion dialogues; and (3) Used the user model to optimize the choice of moves in the persuasion dialogues. For this, we have used

the epistemic approach to probabilistic argumentation. This contrasts with the constellations approach (e.g. [7,12,19]) which is concerned with the uncertainty about the structure of the graph rather than belief in arguments.

The proposal in this paper relies on a user model. This can be generated by querying the user, or by learning from previous interactions with similar users. Some recent studies indicate the potential viability of an empirical approach [5,24].

Most proposals for dialogical argumentation focus on protocols (e.g. [4,8,21,22]). Some strategies have been investigated (e.g. [3,9,18,26]) but the important issue of uncertainty is under-developed. A probabilistic model of the opponent has been used in a dialogue strategy allowing the selection of moves for an agent based on what it believes the other agent is aware of [23]. The history of previous dialogues is used to predict the arguments that an opponent might put forward [10]. For modelling dialogues, a probabilistic finite state machine can represent the possible moves that each agent can make in each state of the dialogue [15]. This has been generalized to POMDPs when there is uncertainty about what an opponent is aware of [11]. However, none of these proposals consider the beliefs of the opposing agent or asymmetric dialogues. In [2], a probabilistic model of persuadee beliefs is used by the persuader to optimize choice of beliefs to present, but there is no consideration of how to get beliefs from the persuadee or how to update the model based on the dialogue. Therefore, the proposal in this paper is an important contribution towards the theoretical foundations for using argumentation in apps for helping persuade users to change behaviour (e.g. eat less, exercise more, drive more carefully, etc.).

Acknowledgements. This research was partly funded by EPSRC grant EP/N008294/1 for the Framework for Computational Persuasion project.

References

1. Baroni, P., Giacomin, M., Vicig, P.: On rationality conditions for epistemic probabilities in abstract argumentation. In: Computational Models of Argument (COMMA 2014), pp. 121–132 (2014)
2. Black, E., Coles, A., Bernardini, S.: Automated planning of simple persuasion dialogues. In: Bulling, N., van der Torre, L., Villata, S., Jamroga, W., Vasconcelos, W. (eds.) CLIMA 2014. LNCS, vol. 8624, pp. 87–104. Springer, Heidelberg (2014)
3. Black, E., Hunter, A.: Reasons and options for updating an opponent model in persuasion dialogues. In: Proceedings of the International Workshop on the Theory and Applications of Formal Argumentation (TAFA 2015) (2015)
4. Caminada, M., Podlaszewski, M.: Grounded semantics as persuasion dialogue. In: Computational Models of Argument (COMMA 2012), pp. 478–485 (2012)
5. Cerutti, F., Tintarev, N., Oren, N.: Formal arguments, preferences, and natural language interfaces to hhuman: an empirical evaluation. In: Proceedings of ECAI 2014, pp. 207–212 (2014)
6. Dung, P.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. Artif. Intell. **77**, 321–357 (1995)

7. Dung, P., Thang, P.: Towards (probabilistic) argumentation for jury-based dispute resolution. In: Computational Models of Argument (COMMA 2010), pp. 171–182. IOS Press (2010)
8. Fan, X., Toni, F.: Assumption-based argumentation dialogues. In: Proceedings of IJCAI 2011, pp. 198–203 (2011)
9. Fan, X., Toni, F.: A general framework for sound assumption-based argumentation dialogues. Artif. Intell. **216**, 20–54 (2014)
10. Hadjinikolis, C., Siantos, Y., Modgil, S., Black, E., McBurney, P.: Opponent modelling in persuasion dialogues. In: Proceedings of IJCAI 2013, pp. 164–170 (2013)
11. Hadoux, E., Beynier, A., Maudet, N., Weng, P., Hunter, A.: Optimization of probabilistic argumentation with Markov decision models. In: Proceedings of IJCAI 2015 (2015)
12. Hunter, A.: Some foundations for probabilistic argumentation. In: Computational Models of Argument (COMMA 2012), pp. 117–128 (2012)
13. Hunter, A.: A probabilistic approach to modelling uncertain logical arguments. Int. J. Approx. Reason. **54**(1), 47–81 (2013)
14. Hunter, A.: Opportunities for argument-centric persuasion in behaviour change. In: Fermé, E., Leite, J. (eds.) JELIA 2014. LNCS, vol. 8761, pp. 48–61. Springer, Heidelberg (2014)
15. Hunter, A.: Probabilistic strategies in dialogical argumentation. In: Straccia, U., Calì, A. (eds.) SUM 2014. LNCS, vol. 8720, pp. 190–202. Springer, Heidelberg (2014)
16. Hunter, A.: Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In: Proceedings of IJCAI 2015 (2015)
17. Hunter, A., Thimm, M.: Probabilistic argumentation with incomplete information. In: Proceedings of ECAI 2014, pp. 1033–1034 (2014)
18. Kontarinis, D., Bonzon, E., Maudet, N., Moraitis, P.: Empirical evaluation of strategies for multiparty argumentative debates. In: Bulling, N., van der Torre, L., Villata, S., Jamroga, W., Vasconcelos, W. (eds.) CLIMA 2014. LNCS, vol. 8624, pp. 105–122. Springer, Heidelberg (2014)
19. Li, H., Oren, N., Norman, T.J.: Probabilistic argumentation frameworks. In: Modgil, S., Oren, N., Toni, F. (eds.) TAFA 2011. LNCS, vol. 7132, pp. 1–16. Springer, Heidelberg (2012)
20. Likert, R.: A technique for the measurement of attitudes. Arch. Psychol. **140**, 1–55 (1932)
21. Prakken, H.: Coherence and flexibility in dialogue games for argumentation. J. Logic Comput. **15**(6), 1009–1040 (2005)
22. Prakken, H.: Formal sytems for persuasion dialogue. Knowl. Eng. Rev. **21**(2), 163–188 (2006)
23. Rienstra, T., Thimm, M., Oren, N.: Opponent models with uncertainty for strategic argumentation. In: Proceedings of IJCAI 2013, pp. 332–338 (2013)
24. Rosenfeld, A., Kraus, S.: Providing arguments in discussions based on the prediction of human argumentative behavior. In: Proceedings of AAAI 2015 (2015)
25. Thimm, M.: A probabilistic semantics for abstract argumentation. In: Proceedings of ECAI 2012, pp. 750–755 (2012)
26. Thimm, M.: Strategic argumentation in multi-agent systems. Kunstliche Intell. **28**(3), 159–168 (2014)