ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-150

Two Dimensional Uncertainty in Persuadee Modelling in Argumentation

Anthony Hunter[1]

Abstract. When attempting to persuade an agent to believe (or disbelieve) an argument, it can be advantageous for the persuader to have a model of the persuadee. Models have been proposed for taking account of what arguments the persuadee believes and these can be used in a strategy for persuasion. However, there can be uncertainty as to the accuracy of such models. To address this issue, this paper introduces a two-dimensional model that accounts for the uncertainty of belief by a persuadee and for the confidence in that uncertainty evaluation. This gives a better modeling for using lotteries so that the outcomes involve statements about what the user believes/disbelieves, and the confidence value is the degree to which the user does indeed hold those outcomes (and this is a more refined and more natural modeling than found in [19]). This framework is also extended with a modelling of the risk of disengagement by the persuadee.

1 INTRODUCTION

Computational models of argument can potentially be used for systems to persuade users to change their behaviour (e.g. to eat less, to exercise more, to use less electricity, to vote, etc) [17]. However, most proposals for dialogical argumentation focus on protocols (e.g. [26, 27, 12, 7]) with strategies being under-developed. See [35] for a review of strategies in multi-agent argumentation.

There are some proposals for using probability theory in dialogical argumentation: A probabilistic model of the opponent is used for selection of moves by an agent based on what it believes the other agent is aware of [31]; The history of previous dialogues is used to predict the arguments that an opponent might put forward [13]; A probabilistic finite state machine can represent the possible moves that each agent can make in each state [18], and generalized to POMDPs when there is uncertainty about what an opponent is aware of [14]. However, none of these use the beliefs of the persuadee or use asymmetric dialogues where only the persuader presents arguments (a requirement when the persuader is a software agent and it is not possible for it to understand natural language arguments from the persuadee). In [4], a probabilistic model of beliefs of the persuadee is used by the persuader to choose beliefs to present, but there is no consideration of update of the model resulting from dialogue, of confidence in the model, of persuasion outcomes involving statements about belief, of expected utility, or of risk of disengagement (which are issues we consider here).

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 [19]. In this paper, that proposal is generalized by introducing a two-dimensional notion of uncertainty with multiple user models and a measure of confidence in them. We will use a logical language to represent and reason with the beliefs in the user models and the confidence in them. This enables a more accurate modelling of expected utility than in [19] since belief statements in the logical language are outcomes in the utility analysis. This is extended with a modelling of risk of disengagement by the persuadee (a key problem when a dialogue is too long) and use this for selecting optimal dialogues.

[1] Department of Computer Science, University College London, London, UK

2 PRELIMINARIES

This paper is based on abstract argumentation [10]. The dialogues concern an argument graph G without self-attacks where $\mathsf{Args}(G)$ is the set of arguments in G, and $\mathsf{Attacks}(G)$ is the set of attack relations in G.

A **system** (the *persuader* running as an app) has a dialogue with a **user** (the *persuadee* using the app) to persuade him/her to believe (or disbelieve) some combination of arguments (e.g. about doing more excercise) as explained in Section 4. The system is aware of all the arguments in the argument graph G whereas the user is not necessarily aware of all the arguments in G.

A **dialogue** is a sequence of moves $D = [m_1, \ldots, m_k]$. Equivalently, we can use D as a function with an index position i to return the move at that index (i.e. $D(i) = m_i$). A **protocol** specifies what moves should/can follow each move in a dialogue.

In this paper, we consider one protocol as an illustration. The only moves are posit of an argument A by the system, denoted $A!$, or termination by the system, denoted $\oplus$, or by the user, denoted $\otimes$. Once terminated, no further moves are possible. An example of unterminated dialogue is $[\mathtt{A}!, \mathtt{C}!, \mathtt{D}!, \mathtt{A}!, \mathtt{C}!, \mathtt{D}!, \mathtt{A}!]$, of a system-terminated dialogue is $[\mathtt{A}!, \mathtt{C}!, \oplus]$, and of a user-terminated dialogue is $[\mathtt{A}!, \mathtt{C}!, \otimes]$.

3 PROBABILISTIC USER MODELS

We will use the epistemic approach to probabilistic argumentation [34, 16, 21, 2].

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

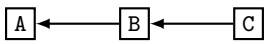
For a mass distribution P, and $A \in \mathsf{Args}(G)$, $P(A)$ is the belief that an agent has in A (i.e. the degree to which the agent believes the premises and the conclusion drawn from those premises). When $P(A) > 0.5$, then the agent believes the argument to some degree,

whereas when $P(A) \leq 0.5$, then the agent disbelieves the argument to some degree.

The following constraint ensures that the mass distribution respects the structure of the graph, without forcing an unattacked argument to be believed [16].

Definition 2. *A mass distribution P is* **rational** *for G iff $\forall (A, B) \in \mathsf{Attacks}(G)$, if $P(A) > 0.5$, then $P(B) \leq 0.5$.*

Example 1. *Consider the following argument graph. Mass distribution $P_1(\mathtt{A}) = 0.6$, $P_1(\mathtt{B}) = 0.9$, and $P_1(\mathtt{C}) = 0.9$ is not rational, whereas $P_2(\mathtt{A}) = 0.6$, $P_2(\mathtt{B}) = 0.3$, and $P_2(\mathtt{C}) = 0.9$ is rational, and $P_3(\mathtt{A}) = 0$, $P_3(\mathtt{B}) = 1$, and $P_3(\mathtt{C}) = 0.3$ is rational.*



The system (the persuader) uses a mass distribution P as a model of the user (the persuadee), and it can update the model at each stage of the dialogue (see Section 7). This is useful for asymmetric dialogues where the user is not allowed to posit arguments. So the only way the user can treat arguments that s/he does not accept is by disbelieving them (and the system aims to reflect this in the user model). In contrast, in symmetric dialogues, the user may be allowed to posit counterarguments to an argument that s/he does not accept.

4 PERSUASION OBJECTIVES

An **objective** is a Boolean combination of arguments. If $A \in \mathsf{Arg}(G)$, then A is a positive literal, and $\neg A$ is a negative literal. Let $\mathsf{AFormulae}(G)$ denote all the formulae that can be formed from the arguments in G using $\wedge$, $\vee$, and $\neg$ as connectives in the usual way.

Informally, an objective is positive or negative from the point of view of the persuader. If it is positive (respectively negative), then the persuader wants the objective to be satisfied (respectively not satisfied) by the arguments believed by the persuader. We consider how to specify whether an objective is positive or negative in Section 9.

In order to formalize the satisfaction of objectives, we treat each subset of $\mathsf{Args}(G)$ as a model (i.e. a possible world).

Definition 3. *The* **satisfaction relation**, *denoted $\models$, is defined as follows where $X \subseteq \mathsf{Args}(G)$, $A \in \mathsf{Args}(G)$, and $\alpha, \beta \in \mathsf{AFormulae}(G)$: (1) $X \models A$ when $A \in X$; (2) $X \models \alpha \wedge \beta$ iff $X \models \alpha$ and $X \models \beta$; (3) $X \models \alpha \vee \beta$ iff $X \models \alpha$ or $X \models \beta$; and (4) $X \models \neg\alpha$ iff $X \not\models \alpha$.*

Essentially $\models$ is a classical satisfaction relation. So if α is a classical tautology, then $X \models \alpha$ for all $X \subseteq \mathsf{Args}(G)$, and if α is a classical contradiction, then $X \not\models \alpha$ for all $X \subseteq \mathsf{Args}(G)$. For $\alpha \in \mathsf{AFormulae}(G)$, let $\mathsf{Models}(\alpha) = \{X \subseteq \mathsf{Args}(G) \mid X \models \alpha\}$. For each graph G, we assume an ordering over the arguments $\langle A_1, ..., A_n \rangle$ so that we can encode each model by a binary number: For a model X, if the ith argument is in X, then the ith digit is 1, otherwise it is 0. E.g. for $\langle \mathtt{A}, \mathtt{B}, \mathtt{C} \rangle$, the model $\{\mathtt{A}, \mathtt{C}\}$ is represented by 101.

According to the user model, the probability of an objective ϕ is the sum of the probability of each model satisfying the objective.

Definition 4. *For $P \in \mathsf{Dist}(G)$, the* **probability of objective** *$\phi \in \mathsf{AFormulae}(G)$ is $P(\phi) = \sum_{X \in \mathsf{Models}(\phi)} P(X)$.*

Suppose $\alpha \in \mathsf{AFormulae}(G)$ and P is a mass distribution. If α is a contradiction of classical logic, then $P(\alpha) = 0$, and if α is a tautology of classical logic, then $P(\alpha) = 1$. Also, if $\{\alpha\} \vdash \beta$, then $P(\alpha) \leq P(\beta)$, and if $\neg(\alpha \wedge \beta)$ is a classical tautology, then $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$.

5 BELIEF STATEMENTS

We use statements (defined next) involving a mass distribution applied to an objective as atoms in a language. These represent the belief a persuadee has in an objective.

Definition 5. *A* **belief statement** *is of the form $P(\alpha) \# x$ where $\alpha \in \mathsf{AFormulae}(G)$ is an objective, $\# \in \{=, \geq, \leq, >, <\}$, and $x \in [0, 1]$. A* **belief formula** *is a Boolean combination of belief statements (i.e. if ϕ is a belief statement, then it is a belief formula, and if ϕ and ψ are belief formulae, then each of $\phi \wedge \psi$, $\phi \vee \psi$ and $\neg\phi$ is a belief formula). Let $\mathsf{BFormulae}(G)$ be the set of belief formulae.*

Example 2. *For $\mathtt{A}, \mathtt{B} \in \mathsf{Args}(G)$, $(P(\mathtt{A} \wedge \mathtt{B}) > 0.9) \vee (P(\neg\mathtt{A} \wedge \neg\mathtt{B}) < 0.5)$ is an example of a belief formula.*

We assume equivalences, denoted $\equiv$, between belief formulae: (1) $P(\alpha) \geq x \equiv (P(\alpha) = x) \vee (P(\alpha) > x)$, (2) $P(\alpha) \leq x \equiv (P(\alpha) = x) \vee (P(\alpha) < x)$, (3) $P(\alpha) \neq x \equiv \neg(P(\alpha) = x)$, (4) $P(\alpha) \not> x \equiv \neg(P(\alpha) > x)$, and (5) $P(\alpha) \not< x \equiv \neg(P(\alpha) < x)$.

Definition 6. *The* **satisfying distributions** *for a belief statement $P(\alpha) \# x$ is $\mathsf{Sat}(P(\alpha) \# x) = \{P' \in \mathsf{Dist}(G) \mid P'(\alpha) \# x\}$, where $\# \in \{=, \geq, \leq, >, <\}$. The set of satisfying distributions for a belief formula is as follows where ϕ and ψ are belief formulae: (1) $\mathsf{Sat}(\phi \wedge \psi) = \mathsf{Sat}(\phi) \cap \mathsf{Sat}(\psi)$; (2) $\mathsf{Sat}(\phi \vee \psi) = \mathsf{Sat}(\phi) \cup \mathsf{Sat}(\psi)$; and (3) $\mathsf{Sat}(\neg\phi) = \mathsf{Sat}(\top) \setminus \mathsf{Sat}(\phi)$.*

Example 3. *For $\langle \mathtt{A}, \mathtt{B} \rangle$, if $P_1(11) = 1$ and $P_2(00) = 1$, then $P_1, P_2 \in \mathsf{Sat}(P(\mathtt{A} \wedge \mathtt{B}) = 1) \vee P(\neg\mathtt{A} \wedge \neg\mathtt{B}) = 1)$. For $\langle \mathtt{C} \rangle$, if $P_3(1) = 0.5$ and $P_4(1) = 0.6$, then $P_3 \notin \mathsf{Sat}(P(\mathtt{C}) > 0.5)$ and $P_4 \in \mathsf{Sat}(P(\mathtt{C}) > 0.5)$.*

Proposition 1. *(1) For $x \in (0, 1]$, $\mathsf{Sat}(P(\bot) = x) = \emptyset$. (2) $\mathsf{Sat}(P(\top) = 1) = \mathsf{Dist}(G)$. (3) For any objective α, $\mathsf{Sat}(P(\alpha) \leq 1) = \mathsf{Dist}(G)$ and $\mathsf{Sat}(P(\alpha) \geq 0) = \mathsf{Dist}(G)$. (4) When $x \neq y$, $\mathsf{Sat}(P(\alpha) = x \wedge P(\alpha) = y) = \emptyset$. (5) When $\vdash \alpha \leftrightarrow \beta$, $\mathsf{Sat}(P(\alpha) = x) = \mathsf{Sat}(P(\beta) = x)$.*

Definition 7. *$\phi, \psi \in \mathsf{BFormulae}(G)$ are* **disjoint** *iff $\mathsf{Sat}(\phi) \cap \mathsf{Sat}(\psi) = \emptyset$.*

Example 4. *Each pair of statements is disjoint: (1) $P(\mathtt{A}) = 0.5, P(\mathtt{A}) = 0.7$; (2) $P(\mathtt{A}) \geq 0.6, P(\mathtt{A}) < 0.5$; (3) $P(\mathtt{A}) > 0.5, P(\neg\mathtt{A}) > 0.7$; and (4) $P(\mathtt{A}) = 0.3, P(\mathtt{A} \wedge \mathtt{B}) = 0.7$.*

Definition 8. *$\{\phi_1, \ldots, \phi_k\} \subseteq \mathsf{BFormulae}(G)$ are* **exhaustive** *iff $\mathsf{Sat}(\phi_1) \cup \ldots \cup \mathsf{Sat}(\phi_k) = \mathsf{Dist}(G)$.*

Example 5. *The set of belief formulae $\{P(\mathtt{A}) > 0.8, P(\mathtt{A}) \leq 0.8 \wedge P(\mathtt{A}) > 0.6, P(\mathtt{A}) \leq 0.6\}$ is exhaustive.*

Proposition 2. *Let $S \subset \mathsf{BFormulae}(G)$ and let $\phi, \phi' \in \mathsf{BFormulae}(G)$. If $S \cup \{\phi\}$ is exhaustive and pairwise disjoint, and $\mathsf{Sat}(\phi) = \mathsf{Sat}(\phi')$, then $S \cup \{\phi'\}$ is exhaustive and pairwise disjoint.*

As we see in Section 9, belief formulae can represent desired/undesired outcomes of a dialogue. The system may want to persuade the user to believe α to some degree (i.e. it is a positive objective). For instance, the system may want the user to believe α above a threshold of 0.9 (i.e. $P(\alpha) > 0.9$). Or it may want to persuade the user to disbelieve α and so α is a negative objective (e.g. $P(\alpha) \leq 0.5$). So the aim of the dialogue is to change the belief/disbelieve in an objective depending on whether it is a positive or negative objective.

6 CONFIDENCE IN BELIEF FORMULAE

The confidence distribution is a probability distribution over mass distributions. It gives the probability that a given user model is the correct representation of the user's beliefs.

Definition 9. *A* **confidence distribution** *is* $Pr : \mathsf{Dist}(G) \rightarrow [0,1]$ *s.t.* $\sum_{P \in \mathsf{Dist}(G)} Pr(P) = 1$. *For a belief formula* ϕ, *a* **formula confidence** *is* $Pr(\phi) = \sum_{P \in \mathsf{Sat}(\phi)} Pr(P)$.

For instance, the formula confidence $Pr(P(\mathtt{A}) = 0.7) > 0.5$ means that the persuader has at least 0.5 confidence in the persuadee belief in A being 0.7.

Example 6. *For* $\langle \mathtt{A}, \mathtt{B} \rangle$, *consider* P_1, P_2, *and* P_3, *defined below. Some examples of confidence are:* $Pr(P(\mathtt{A}) = 1) = 1/2$, $Pr(P(\mathtt{A}) \geq 1/2) = 1$, $Pr(P(\mathtt{A}) = 1/2) = 1/2$, $Pr(P(\mathtt{B}) = 0) = 1/4$, $Pr(P(\neg\mathtt{B}) = 1) = 1/4$, $Pr(P(\neg\mathtt{B}) = 1/2) = 1/4$, *and* $Pr(P(\mathtt{A} \wedge \mathtt{B}) \geq 1/4) = 3/4$.

	$Pr(P_1) = 1/2$	$Pr(P_2) = 1/4$	$Pr(P_3) = 1/4$
11	*1*	*0*	*1/4*
10	*0*	*1/2*	*1/4*
01	*0*	*0*	*1/4*
00	*0*	*1/2*	*1/4*

Clearly, for all Pr, $Pr(P(\bot) = 0) = 1$, $Pr(P(\bot) = 1) = 0$, $Pr(P(\top) = 1) = 1$, and $Pr(P(\top) = 0) = 0$.

Proposition 3. *For objectives* α, *and* β, *and* $x, y, z \in [0,1]$, *formula confidence satisfies: (1)* $Pr(P(\alpha) \geq x) > z$ *if* $Pr(P(\alpha) \geq y) > z$ *and* $y \geq x$; *(2)* $Pr(P(\alpha) \geq x) \geq Pr(P(\alpha) \geq y)$ *when* $y \geq x$; *(3)* $Pr(P(\alpha) \geq x) \geq Pr(P(\beta) \geq x)$ *where* $\{\beta\} \vdash \alpha$; *(4)* $Pr(P(\alpha) \geq x) = Pr(P(\neg\alpha) \leq (1-x))$; *(5)* $Pr(P(\alpha) \geq 0.5 \vee P(\beta) \geq 0.5) = 1$ *where* $\{\beta\} \vdash \neg\alpha$; *(6)* $Pr(P(\alpha) \geq x) = Pr(P(\alpha) = x) + Pr(P(\alpha) > x)$; *and (7)* $Pr(P(\alpha) < x) + Pr(P(\alpha) = x) + Pr(P(\alpha) > x) = 1$.

If there is positive confidence that the attacker (respectively attackee) is believed, then there is positive confidence that the attackee (respectively attacker) is not believed.

Proposition 4. *Let* Pr *be s.t. if* $Pr(P') > 0$, *then* P' *is rational. For all* $(A, B) \in \mathsf{Attacks}(G)$,

1. *if* $Pr(P(A) > 0.5) > 0.5$, *then* $Pr(P(B) \leq 0.5) > 0.5$.
2. *if* $Pr(P(B) > 0.5) > 0.5$, *then* $Pr(P(A) \leq 0.5) > 0.5$.

The following results ensure that we can use belief formulae as outcomes in a lottery (Section 10).

Proposition 5. *If* $\{\phi_1, \ldots, \phi_n\} \subseteq \mathsf{BFormulae}(G)$ *is exhaustive, then* $Pr(\phi_1 \vee \ldots \vee \phi_n) = 1$.

Proposition 6. *If* $\phi, \psi \in \mathsf{BFormulae}(G)$ *are disjoint, then* $Pr(\phi \vee \psi) = Pr(\phi) + Pr(\psi)$.

We can treat atoms in $\mathsf{BFormulae}(G)$ as atoms in a classical propositional language, thereby use $\vdash$ as the classical propositional consequence relation.

Proposition 7. *Let* $\phi, \psi \in \mathsf{BFormulae}(G)$. *If* $\{\phi\} \vdash \psi$, *then* $Pr(\phi) \leq Pr(\psi)$.

As the number of different mass distributions with non-zero confidence increases, the confidence in some belief formulae will fall.

Proposition 8. *Let* Pr^1 *and* Pr^2 *be confidence distributions, and let* $\mathsf{Dom}(Pr) = \{P \mid Pr(P) > 0\}$. *If* $\mathsf{Dom}(Pr^1) \subseteq \mathsf{Dom}(Pr^2)$, *then there is a* $\phi \in \mathsf{BFormulae}(G)$ *such that* $Pr^1(\phi) \geq Pr^2(\phi)$.

The confidence value is important for two reasons. First, it gives a better modeling for using lotteries so that the outcomes involve statements about what the user believes/disbelieves, and the confidence value is the degree to which the user does indeed hold those outcomes (and this is a more refined and more natural modeling than found in [19]). Second, it allows for uncertainty about the user to be better managed. If we are sure we know what the user believes, then have one probability distribution, whereas for example, if we are not sure about the user we have, we may have multiple distributions.

So we will treat belief statements as outcomes in a lottery (for calculating expected utility), and use a confidence distribution to give the probability that we get that outcome.

7 UPDATING USER MODELS

This section reviews some proposals in [19]. To update a user model during a dialogue, a mass redistribution function takes a mass distribution and returns a revised mass distribution. Possibilities for this include probabilistic conditioning. However, in this paper, we use an alternative defined next for redistributing mass from models (i.e. possible worlds) not satisfying α to models satisfying α.

Definition 10. *[19] Let* $\alpha \in \mathsf{AFormulae}(G)$ *be a literal, let* P *be a mass distribution, and let* $k \in [0,1]$. *A* **refinement function**, *denoted* $H^k_\alpha(P)$, *returns the mass distribution* P' *as follows where* $X \in \mathsf{Models}(G)$

$$P'(X) = \begin{cases} P(X) + (k \times P(h_\alpha(X))) & \text{if } X \models \alpha \\ (1-k) \times P(X) & \text{if } X \not\models \alpha \end{cases}$$

and where $h_\alpha(X) = X \setminus \{A\}$ *when* α *is of the form* A *and* $h_\alpha(X) = X \cup \{A\}$ *when* α *is of the form* $\neg A$.

The above function is called refinement because it refines the mass distribution using an update. See Table 1 for examples of redistribution using the refinement function. In the above definition, h_α returns the model closest to X but with α no longer satisfied. If $k = 1$, then all the mass is transferred from the models not satisfying α to models satisfying α. If $k < 1$, then only a proportion is transferred. This gives flexibility to model update in different kinds of user. For instance, if we want to model a user that when conceding an argument is believable, s/he does not fully believe the argument, we can use $k < 1$ to update the model so that the argument is not fully believed in the model.

Table 1. Examples of mass redistribution

AB	P	$H^1_A(P)$	$H^1_{\neg A}(P)$	$H^{0.75}_A(P)$	$H^1_B(P)$
11	0.6	0.7	0.0	0.675	0.8
10	0.2	0.3	0.0	0.275	0.0
01	0.1	0.0	0.7	0.025	0.2
00	0.1	0.0	0.3	0.025	0.0

Given a mass distribution P, representing a user's beliefs at the current state of the dialogue, we want to update the model depending on the move made. For this, we consider the notion of an update method $\sigma(P_{i-1}, D(i)) = P_i$ which generates a mass distribution P_i

from P_{i-1} based on the move $D(i)$. Each method σ is defined as a rule with a condition (based on the move, the current mass distribution, and the graph), and a consequent that specifies the redistribution.

To illustrate, the trusting method (below) raises the belief in a posit, and lowers the belief in attackers and attackees.

Definition 11. *[19] For step i in the dialogue, the* **trusting method** *generates P_i from P_{i-1} as follows, where $\Phi = \{\neg C \mid (A, C) \in \mathsf{Attacks}(G)$ or $(C, A) \in \mathsf{Attacks}(G)\}$.*

$$\text{If } D(i) = A!, \text{ then } P_i = H^1_\Phi(H^1_A(P_{i-1})).$$

Example 7. *For $\langle \mathtt{A}, \mathtt{B}\rangle$, consider the argument graph in Example 1 with dialogue $[\mathtt{A}!, \oplus]$ and the trusting method. Let the initial mass be $P_0(011) = 0.3$, $P_0(010) = 0.2$, $P_0(001) = 0.3$, and $P_0(000) = 0.2$. After move $\mathtt{A}!$, $P_1(101) = 0.6$, and $P_1(100) = 0.4$.*

The strict method (defined next) only allows a posit to update the belief in the posit when there is no attacker of the posit that is believed.

Definition 12. *[19] For step i in the dialogue, the* **strict method** *generates P_i from P_{i-1} as follows, where $\Phi = \{\neg C \mid (A, C) \in \mathsf{Attacks}(G)\}$.*

$$\begin{array}{l}\text{If } D(i) = A!,\\ \text{and for all } (B, A) \in \mathsf{Attacks}(G), P_{i-1}(B) \leq 0.5,\\ \text{then } P_i = H^1_\Phi(H^1_A(P_{i-1})), \text{ else } P_i = P_{i-1}\end{array}$$

Example 8. *For $\langle \mathtt{A}, \mathtt{B}, \mathtt{C}\rangle$, consider the graph in Example 1 with dialogue $[\mathtt{A}!, \mathtt{C}!, \mathtt{A}!, \oplus]$ and the strict method. Let the initial mass be $P_0(111) = 0.2$, $P_0(110) = 0.3$, $P_0(011) = 0.3$, and $P_0(010) = 0.2$. After the first $\mathtt{A}!$, $P_1(111) = 0.2$, $P_1(110) = 0.3$, $P_1(011) = 0.3$, and $P_1(010) = 0.2$. After $\mathtt{C}!$, $P_2(101) = 0.5$, and $P_2(001) = 0.5$. After the second $\mathtt{A}!$, $P_3(101) = 1$.*

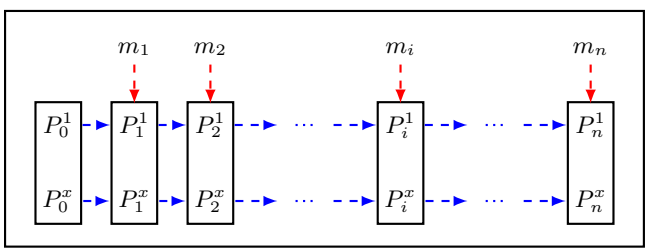


Figure 1. Schematic of the update of the 2D model where $D = [m_1, \ldots, m_n]$ and $\mathcal{P}_0 = \langle P^1_0, \ldots, P^x_0\rangle$. At the end of the dialogue, the user models are $\mathcal{P}_n = \langle P^1_n, \ldots, P^x_n\rangle$.

See [19] for further update methods and for more discussion of how they are used. These are only illustrative of updates methods. With a wider range of moves, a wider range of update methods can be considered. For instance, with moves to get information from the user, further update methods can be defined.

8 2D MODELS

We combine user models (i.e. a mass distribution representing the persuadee beliefs) with the confidence distribution.

Definition 13. *A* **2D model** *is a tuple $(\mathcal{P}, Pr)$ where $\mathcal{P}$ is a tuple $\langle P^1, \ldots, P^x\rangle$ s.t. each P^i in $\mathcal{P}$ is a mass distribution and Pr is a confidence distribution s.t. $\sum_{i=1}^{x} Pr(P^i) = 1$.*

So each P^i in the tuple denotes a mass distribution modelling the user. We may have different ones because we are unsure which is correct, though some may be identical.

The most certain 2D model is when the mass distributions are identical (i.e. for all $P^i, P^j \in \mathcal{P}$, if $Pr(P^i) > 0$ and $Pr(P^j) > 0$, then $P^i \equiv P^j$).

At the other extreme, the least certain 2D model is when there is a $P \in \mathcal{P}$ for each $X \subseteq \mathsf{Args}(G)$ such that $P(X) = 1$, and where $Pr(P) = 1/k$ s.t. $k = 2^n$ and $|\mathsf{Arg}(G)| = n$.

In the following definition for updating the 2D model, we can use a different update method for each user model to mimic different ways a user might update his/her beliefs.

Definition 14. *Let $(\mathcal{P}_{i-1}, Pr_{i-1})$ is a 2D model where $\mathcal{P}_{i-1} = \langle P^1_{i-1}, \ldots, P^x_{i-1}\rangle$, let m_i be a move, and let σ^j be the update method for user model P^j. The* **2D model update** *$\mathcal{P}_{i+1}$ is $\langle \sigma^1(P^1_{i-1}, m_i), \ldots, \sigma^x(P^x_{i-1}, m_i)\rangle$ where for all $P^j \in \mathcal{P}_i$, $Pr_i(P^j_i) = Pr_{i-1}(P^j_{i-1})$.*

So for each step i of the dialogue, the above definition updates P_{i-1} to give P_i. This is represented schematically in Figure 1. For a dialogue D with n moves, and initial 2D model $(\mathcal{P}_0, Pr_0)$ s.t. $\mathcal{P}_0 = \langle P^1_0, \ldots, P^x_0\rangle$, we use the function $\mathsf{Update}(\mathcal{P}_0, Pr_0, D) = (\mathcal{P}_n, Pr_n)$ to denote the iterative application of the above definition starting with m_1, then m_2, and so on, until m_n.

Example 9. *For $\langle \mathtt{A}, \mathtt{B}, \mathtt{C}\rangle$, consider the argument graph in Example 1 with dialogue $D = [\mathtt{C}!, \mathtt{A}!, \oplus]$. Let $\mathcal{P}_0 = \langle P^1_0, P^2_0\rangle$ where $P^1_0(m) = 1/8$ for all models, and $P^2_0(011) = 1/2$ and $P^2_0(010) = 1/2$. Also let σ^1 be strict update and σ^2 be trusting update. For move $m_1 = \mathtt{C}!$, $\mathcal{P}_1 = \langle P^1_1, P^2_1\rangle$ where $P^1_1(101) = 1/2$ and $P^1_1(001) = 1/2$, and $P^2_1(001) = 1$. Then for move $m_2 = \mathtt{A}!$, $\mathcal{P}_2 = \langle P^1_2, P^2_2\rangle$ where $P^1_2(101) = 1$, and $P^2_2(101) = 1$. So $\mathsf{Update}(\mathcal{P}_0, Pr_0, D) = (\mathcal{P}_2, Pr_2)$.*

For some update functions, e.g. the trusting method, and some belief formulae, we can always construct a dialogue that will result in total confidence in the formulae, and so the mass distributions in $\mathcal{P}_0$ become more similar. For instance, the following result shows that for a conflictfree set of arguments, each argument can be posited in a dialogue, and the trusting update method ensures that they are all believed.

Proposition 9. *For a 2D model $(\mathcal{P}_0, Pr_0)$, and belief statement π of the form $P(\alpha) \geq 0.5$, where α is a conjunction of arguments, if σ is the trusting update method for all $P \in \mathcal{P}$, and there are no conjuncts A, B in α such that $(A, B) \in \mathsf{Attacks}(G)$, then there is a dialogue $D = [m_1, \ldots, m_n]$ s.t. $Pr_n(\pi) = 1$, where $\mathsf{Update}(\mathcal{P}_0, Pr_0, D) = (\mathcal{P}_n, Pr_n)$.*

To recap, the 2D model allows us to use multiple user models and multiple update methods to represent the persuadee.

9 UTILITY CONSTRAINTS

The objectives introduced in Section 4 represent what the persuader wants the persuadee to believe or disbelieve.

Definition 15. *An* **objective tuple** *is a pair (Q^+, Q^-) where $Q^+ \subseteq \mathsf{AFormulae}(G)$ and $Q^- \subseteq \mathsf{AFormulae}(G)$ such that $Q^+ \cap Q^- = \emptyset$. We refer to Q^+ as the set of positive objectives and Q^- as the set of negative objectives.*

Example 10. *An example of an objective tuple is* $(\{\mathtt{A}\}, \{\mathtt{B}\})$ *where* A *is "You are doing little excercise, and so you should do a brisk 30min walk everyday" and* B *is "Sugar-loaded sports drinks are advertised for sports people, and therefore they are healthy".*

So a positive objective is an objective that the system wants the user to believe and a negative objective is an objective that the system wants the user to disbelieve. Therefore, for an objective tuple (Q^+, Q^-), outcomes are belief statements as tabulated below. Hence, for a positive objective α, belief in α is a positive outcome, and disbelief in α is a negative outcome. For example, for a positive objective, α, $P(\alpha) > 0.9$ is a positive outcome and $P(\alpha) < 0.4$ is a negative outcome. Similarly, for a negative objective α, belief in α is a negative outcome, and disbelief in α is a positive outcome.

objective	x	belief statement as outcome
α is +ve	$x \in (0.5, 1]$	$P(\alpha) > x$ is +ve outcome.
α is -ve	$x \in (0.5, 1]$	$P(\alpha) > x$ is -ve outcome.
α is +ve	$x \in [0, 0.5]$	$P(\alpha) \leq x$ is -ve outcome.
α is -ve	$x \in [0, 0.5]$	$P(\alpha) \leq x$ is +ve outcome.

We can generalize to arbitary formulae in $\mathsf{BFormulae}(G)$ as follows: If ϕ and ψ are +ve (respectively -ve) outcomes, then $\phi \wedge \psi$ and $\phi \vee \psi$ are +ve (respectively -ve) outcomes. And if ϕ is a +ve (respectively -ve) outcome, then $\neg\phi$ is a -ve (respectively +ve) outcome.

Definition 16. *A* **persuasion utility function***, denoted U, for an objective tuple (Q^+, Q^-) is an assignment from $\mathsf{BFormulae}(G)$ to $\mathbb{R}$ such that: (1) If ϕ is a +ve outcome, then $U(\phi) > 0$; (2) If ϕ is a -ve outcome, then $U(\phi) < 0$.*

Example 11. *Continuing Example 10, we can choose the outcomes and U such that $U(P(\mathtt{A}) > 0.9) = 10$, $U(P(\mathtt{A}) > 0.5 \wedge P(\mathtt{A}) \leq 0.9) = 8$, $U(P(\mathtt{A}) \leq 0.5) = -10$, $U(P(\mathtt{B}) > 0.5) = 5$, and $U(P(\mathtt{B}) \leq 0.5) = -5$.*

Note, if a formulae is neither +ve nor -ve, it is not necessarily of zero utility. For example, let ϕ be +ve, and ψ be -ve, then $U(\phi \wedge \psi)$ might be greater than 0 if ϕ is more important than ψ, or less than 0 if ψ is more important than ϕ.

Definition 17. *A persuasion utility function U for (Q^+, Q^-) is* **sensible** *iff U satisfies the following conditions.*

1. *If $x > y$, and α is a +ve (resp. -ve) objective, then $U(P(\alpha) \geq x) \geq U(P(\alpha) \geq y)$ (resp. $U(P(\alpha) \geq x) \leq U(P(\alpha) \geq y)$).*
2. *If $\{\alpha\} \vdash \beta$, and α, β are +ve (resp. -ve) objectives, then $U(P(\alpha) \geq x) \geq U(P(\beta) \geq x)$ (resp. $U(P(\alpha) \geq x) \leq U(P(\beta) \geq x)$).*
3. *If $\{\phi\} \vdash \psi$, and ϕ, ψ are +ve (resp. -ve) outcomes, then $U(\phi) \geq U(\psi)$ (resp. $U(\phi) \leq U(\psi)$).*

This definition provides intuitive constraints on the persuasion utility function. Condition 1 ensures increased (resp. decreased) belief in a +ve (resp. -ve) objective has increased (resp. decreased) utility; Condition 2 ensures belief in an inferentially stronger +ve (resp. -ve) objective has increased (resp. decreased) utility; and Condition 3 ensures an inferentially stronger +ve (resp. -ve) outcome has increased (resp. decreased) utility.

Proposition 10. *If (Q^+, Q^-) is an objective tuple, then there is a persuasion utility function U for (Q^+, Q^-) such that U is sensible.*

So if the positive and negative objectives are disjoint, then we are guaranteed to identify a persuasion utility function that is sensible (in the sense of Definition 17).

10 EXPECTED UTILITY

A lottery with possible outcomes o_1,..,o_n that are pairwise disjoint and exhaustive (i.e. exactly one of them is guaranteed to occur), that occur with probabilities $p_1, .., p_n$ respectively, is written as $[p_1, o_1; ...; p_n, o_n]$. For a utility function U, the expected utility of a lottery L is $\sum_{i=1}^{n} p_i \times U(o_i)$. We harness this notion of a lottery as follows.

Definition 18. *Let D be a dialogue, let $S = \{\phi_1, \ldots, \phi_k\}$ be a set of disjoint and exhaustive outcomes (i.e. belief formulae), let $(\mathcal{P}_0, Pr_0)$ be the initial 2D model, let $\mathsf{Update}(\mathcal{P}_0, Pr_0, D) = (\mathcal{P}_n, Pr_n)$, and let U be a utility function. The* **lottery** *for Pr, U, S is $\mathsf{Lot}(Pr, U, S) =$*

$$[Pr(\phi_1), \phi_1; \ldots; Pr(\phi_k), \phi_k]$$

Then the **expected utility** *for Pr, U, S is $\mathsf{EU}(Pr, U, S) =$*

$$(Pr(\phi_1) \times U(\phi_1)) + \ldots + (Pr(\phi_k) \times U(\phi_k))$$

Example 12. *For $\langle \mathtt{A}, \mathtt{B} \rangle$, let $P_n^1(11) = 1$, $P_n^2(11) = 0.6$, $P_n^2(01) = 0.4$, and $P_n^3(01) = 1$, with $Pr(P_n^1) = 0.5$, $Pr(P_n^2) = 0.3$, and $Pr(P_n^3) = 0.2$. Let the objective tuple be $(\{\mathtt{A}\}, \emptyset)$. Hence, ϕ_1 is a positive outcome, ϕ_2 is neither a positive nor negative outcome, and ϕ_3 is a negative outcome. So using the values for Pr and U in the table, the expected utility is 4.5.*

ϕ	$Pr(\phi)$	$U(\phi)$
$\phi_1 = P(\mathtt{A}) > 0.9$	*0.5*	*10*
$\phi_2 = (P(\mathtt{A}) \leq 0.9) \wedge (P(\mathtt{A}) > 0.5)$	*0.3*	*5*
$\phi_3 = P(\mathtt{A}) \leq 0.5$	*0.2*	*-10*

Using the 2D model, we can determine the optimal dialogues for a lottery as follows.

Definition 19. *A dialogue D is* **optimal** *w.r.t. the initial 2D model $(\mathcal{P}_0, Pr_0)$, utility function U, and $\mathsf{Update}(\mathcal{P}_0, Pr_0, D) = (\mathcal{P}_n, Pr_n)$, when $\mathsf{EU}(Pr_n, U, S)$ is maximized.*

In the following example, we show how we can choose between dialogues using a 2D model.

Example 13. *Consider the following argument graph with the objective tuple $(\{\mathtt{A} \vee \mathtt{C}\}, \emptyset)$.*

Let $\mathcal{P}_0 = \langle P_0^1, P_0^2 \rangle$ be defined as follows, and assume we use the strict update method. Note, we give the probability for each argument rather than each model to save space.

	A	B	C	D
P_0^1	*0*	*1*	*0*	*0*
P_0^2	*0*	*0*	*0*	*1*

The updated mass distributions are given below for dialogue $D_1 = [\mathtt{A}!, \oplus]$ (left) and for dialogue $D_2 = [\mathtt{C}!, \oplus]$ (right).

	A	B	C	D
P_0^1	*0*	*1*	*0*	*0*
P_0^2	*1*	*0*	*0*	*1*

	A	B	C	D
P_0^1	*0*	*1*	*1*	*0*
P_0^2	*0*	*0*	*0*	*1*

Assume $Pr(P_n^1) = 2/3$ *and* $Pr(P_n^2) = 1/3$. *and outcomes* $\phi_1 = P(\mathtt{A} \vee \mathtt{C}) > 0.5$ *and* $\phi_2 = P(\mathtt{A} \vee \mathtt{C}) \leq 0.5$ *s.t.* $U(\phi_1) = 5$ *and* $U(\phi_2) = -5$. *So for dialogue* D_1, *expected utility is* $(1/3 \times 5) + (2/3 \times -5) = -5/3$, *and for dialogue* D_2, *expected utility is* $(2/3 \times 5) + (1/3 \times -5) = 5/3$.

We could select a longer dialogue $D_3 = [\mathtt{A}!, \mathtt{C}!, \oplus]$ *giving the following updated mass distributions, and with expected utility* $(1 \times 5) + (0 \times -5) = 5$.

	A	B	C	D
P_1	*0*	*1*	*1*	*0*
P_2	*1*	*0*	*0*	*1*

For the shorter dialogues, D_2 *is better than* D_1. *However,* D_3 *is better than both* D_2 *and* D_1, *but* D_3 *is longer.*

At one extreme, if the 2D model only contains one user model, then the outcome is known with certainty (i.e there is complete confidence in the belief statement).

Proposition 11. *If* $[Pr(\phi_1), \phi_1; \ldots; Pr(\phi_k), \phi_k]$ *is a lottery, and* $|\mathcal{P}| = 1$, *then there is a* $\phi_i \in \{\phi_1, \ldots, \phi_k\}$ *s.t.* $Pr(\phi_i) = 1$, *and for all* $\phi_j \in \{\phi_1, \ldots, \phi_k\} \setminus \{\phi_i\}$, $Pr(\phi_i) = 0$.

Example 14. *Consider the disjoint and exhaustive outcomes* $P(\mathtt{A} \vee \mathtt{B}) = 1$, $P(\neg\mathtt{A} \wedge \neg\mathtt{B}) = 1$, *and* $P(\mathtt{A} \vee \mathtt{B}) < 1 \wedge P(\neg\mathtt{A} \wedge \neg\mathtt{B}) < 1$. *Let* $\mathcal{P} = \{P'\}$ *and so* $Pr(P') = 1$. *Let* $P'(11) = 1$. *Hence,* $Pr(P(\mathtt{A} \vee \mathtt{B}) = 1) = 1$.

At the other extreme, there are various situations that give rise to a uniform distribution over the outcomes. We consider the following which reflects the ignorance when there are multiple mass distributions with no agreement.

Proposition 12. *Let* $[Pr(\phi_1), \phi_1; \ldots; Pr(\phi_n), \phi_n]$ *be a lottery where* Pr *is a uniform distribution over* $\mathcal{P}$. *Also for each* $P \in \mathcal{P}$, *there is a* $X \subseteq \mathsf{Args}(G)$ *s.t.* $P(X) = 1$, *and for each* $X \subseteq \mathsf{Args}(G)$, *there is a* $P \in \mathcal{P}$ *s.t.* $P(X) = 1$. *If there is an* $x > 0$ *s.t. for each* $\phi_i \in \{\phi_1, \ldots, \phi_n\}$, $|\mathsf{Sat}(\phi_i)| = x$, *then for each* $\phi_i, \phi_j \in \{\phi_1, \ldots, \phi_n\}$, $Pr(\phi_i) = Pr(\phi_j)$.

Example 15. *For* $\langle \mathtt{A} \rangle$, *let* $\mathcal{P} = \{P_1, P_2\}$ *where* $P_1(1) = 1$ *and* $P_2(0) = 1$. *Let* $Pr(P_1) = 1/2$ *and* $Pr(P_2) = 1/2$. *Consider the outcomes* $P(\mathtt{A}) > 0.5$ *and* $P(\neg\mathtt{A}) > 0.5$. *Hence,* $Pr(P(\mathtt{A}) > 0.5) = 1/2$ *and* $Pr(P(\neg\mathtt{A}) > 0.5) = 1/2$

Between these extremes, the 2D model can be valuable in identifying the optimal dialogues. Note, that normally we do not envisage that the 2D model will contain many mass distributions. Furthermore, we will focus on update methods that are computationally efficient. Hence, we envisage the approach is computationally viable.

11 MODELLING DISENGAGEMENT

For every user-terminated dialogue D, there is a probability that the user of the app will disengage before the end of the dialogue (e.g. through loss of interest). This can rise as the length of the dialogue increases. We assume a **stay-in probability**, denoted q, which for step i in the dialogue is the probability that the user will remain engaged for the next step $i + 1$. We assume no disengagement after the ultimate posit.

Definition 20. *Let* $D = [m_1, \ldots, m_n]$ *be a system-terminated dialogue with stay-in probability* q. *If* $n > 2$, *the* **probability of engagement** *is* $prob_{engage} = q^{n-2}$ *and the* **probability of disengagement** *is* $prob_{disengage} = \sum_{i=1}^{n-2} q^{(i-1)} \times (1 - q)$. *If* $n = 1$ *or* $n = 2$, *then* $prob_{engage}$ *is 1, and* $prob_{disengage}$ *is 0.*

Example 16. *Consider the dialogue in Figure 2 with the stay-in probability being 0.9. So* $prob_{engage}$ *is* $0.9 \times 0.9 = 0.81$ *and* $prob_{disengage}$ *is* $0.1 + (0.9 \times 0.1) = 0.19$.

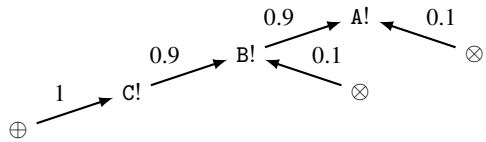


Figure 2. For dialogue $[\mathtt{A}!, \mathtt{B}!, \mathtt{C}!, \oplus]$, each node is a move. The left branch is the system-terminated dialogue, and each branch that ends in $\otimes$ is a user-terminated dialogue. Each arc in the tree is labelled with the probability of engagement (leftwards) or disengagement (rightwards).

Proposition 13. *For a system-terminated dialogue* D, *and a stay-in probability* q, $prob_{engage} + prob_{disengage} = 1$.

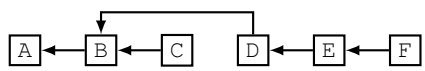
Since disengagement is often a clear event in a dialogue, obtaining a stay-in probability can be obtained from analyzing previous dialogues for a class of users.

Given the probability of engagement $Prob_{engage}$ and a lottery $[Pr(\phi_1), \phi_1; \ldots; Pr(\phi_n), \phi_n]$, we form a revised lottery as specified in the following result where $\otimes$ denotes the outcome of disengagement..

Proposition 14. *If* $[Pr(\phi_1), \phi_1; \ldots; Pr(\phi_n), \phi_n]$ *is a lottery, and* $Prob_{engage} \in [0, 1]$, *then the following is a lottery where for each outcome* ϕ_i, $Pr^*(\phi_i)$ *is* $Pr(\phi_i) \times Prob_{engage}$.

$$[Pr^*(\phi_1), \phi_1; \ldots; Pr^*(\phi_n), \phi_n; 1 - Prob_{engage}, \otimes]$$

Example 17. *For* $\langle \mathtt{A}, \ldots, \mathtt{F} \rangle$, *consider the graph with* $\mathcal{P} = \{P_0\}$ *where* $P_0(010010) = 1$. *So* $P_0(\mathtt{B}) = 1$ *and* $P_0(\mathtt{E}) = 1$.



Let $\mathtt{A}$ *be a +ve objective, and let* $P(\mathtt{A}) \geq 0.9$ *and* $P(\mathtt{A}) < 0.9$ *be outcomes in the lottery. So* $D_1 = [\mathtt{C}!, \mathtt{A}!, \oplus]$ *and* $D_2 = [\mathtt{F}!, \mathtt{D}!, \mathtt{A}!, \oplus]$ *are dialogues that terminate with* $Pr(P(\mathtt{A}) \geq 0.9) = 1$ *according to the strict update method. Let the stay-in probability be 3/4. So* $Prob_{engage} = 3/4$ *for* D_1 *and* $Prob_{engage} = 9/16$ *for* D_2. *Hence, the revised lottery has for* D_1, $Pr^*(P(\mathtt{A}) \geq 0.9) = 3/4$, $Pr^*(P(\mathtt{A}) < 0.9) = 0$, *and* $Pr^*(\otimes) = 1/4$, *and for* D_2, $Pr^*(P(\mathtt{A}) \geq 0.9) = 9/16$, $Pr^*(P(\mathtt{A}) < 0.9) = 0$, *and* $Pr^*(\otimes) = 7/16$. *So for this stay-in probability, the optimal dialogue is* D_1.

	A	B	C	D	E	F
P_0	*0*	*1*	*0*	*0*	*1*	*0*
P_n *for* D_1	*1*	*0*	*1*	*0*	*0*	*0*
P_n *for* D_2	*1*	*0*	*0*	*1*	*0*	*1*

Shorter dialogues can be preferable (as above). In general, we trade a decrease in expected utility for a decrease in risk of disengagement (e.g. for Example 13 whether D_2 or D_3 is optimal would depend on the stay-in probability).

12 DISCUSSION

This paper provides the following contributions (which are potentially important features for using argumentation in software for changing behaviour): (1) A 2D model of uncertainty giving predictions of the beliefs of the persuadee, and of the confidence in those predictions; (2) A framework for updating the 2D model through dialogues; and (3) Shown how the 2D model can be used to optimize choice of moves while taking into account the risk of disengagement.

For this, the epistemic approach to probabilistic argumentation has been used. This contrasts with the constellations approach (e.g. [11, 23, 15]) which is concerned with the uncertainty about the structure of the graph rather than belief in arguments.

The proposal in this paper relies on 2D models. This can be generated by querying the user, or by learning from previous interactions with the user or similar users. Some recent studies indicate the potential viability of an empirical approach [30, 9, 33].

Utility theory has been considered previously in argumentation (for example [29, 32, 24, 25]) though none of these represent the uncertainty of moves made by each agent in argumentation. There is an approach using expected utility where outcomes are specified as particular arguments being included or excluded from extensions [20], but it is based on the constellations approach (as opposed to the epistemic approach), and there is no consideration of updates to the model. Outcomes from asymmetric dialogues have also been considered in [5], but that work focuses on whether it is guaranteed, possible, or impossible to present a winning coalition of arguments with respect to grounded semantics, and there is no consideration of uncertainty.

There is increasing interest in formalizing the notion of the strength of an argument, with a number of proposals (e.g. [3, 8, 24, 22, 1, 6, 28]). It would be interesting to investigate the pros and cons of using these conceptualizations of strength of an argument instead of epistemic probabilities in this framework. Nonetheless, some clear advantages of the epistemic approach are the clear semantics for the evaluation of the arguments, the ease with which epistemic approach can be used in a lottery, and the possibility to obtain the probabilities by analysing statistical data concerning the behaviour agents.

The work in this paper goes beyond [19]: (1) to better model lotteries so that outcomes involve statements about user beliefs, and the confidence value is the degree to which the user does indeed hold those outcomes which is a more refined and natural modeling than [19]; (2) to allow for uncertainty about the user to be handled, and so if we are sure we know the users beliefs, then we have one distribution, whereas if we are unsure about kind of user, we have multiple distributions; and (3) to model the risk of disengagement which is a practical issue that significantly affects the usability of any argumentation approach for behavior change.

Our current research is directed at generating probability distributions for user models. We are exploring the use of queries to the user where the user can express belief in individual arguments (such as strongly agree, agree, neither agree nor disagree, etc which are then mapped to the [0,1] interval). If we do this for some arguments, we can attempt to guess the belief in remaining arguments. We are also exploring how classes of user might believe/disbelieve certain arguments. So by knowing beliefs in arguments for some members of the class, and by having criteria for assigning individuals to a class with some probability, we may construct the 2D model for a user. We envisage that by surveying representative samples of individuals, we can obtain useful 2D models. We aim to develop similar methods to those used in [30, 9, 33]. We plan to undertake empirical evaluation of the approach in apps to persuade users to change their behaviour with respect to some aspect of their lifestyle (e.g. to eat less, to drink less alcohol, to drive more safely, to recycle more, etc). We see the theoretical developments in this paper being viable and valuable for the prototype system that we are implementing.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] L. Amgoud and J. Ben-Naim, 'Axiomatic foundations of acceptability semantics', in *Proceedings of KR'16*, pp. 2–11, (2016).
[2] P. Baroni, M. Giacomin, and P. Vicig, 'On rationality conditions for epistemic probabilities in abstract argumentation', in *Computational Models of Argument (COMMA'14)*, pp. 121–132, (2014).
[3] Ph. Besnard and A. Hunter, 'A logic-based theory of deductive arguments', *Artificial Intelligence*, **128**(1-2), 203–235, (2001).
[4] E. Black, A Coles, and S Bernardini, 'Automated planning of simple persuasion dialogues', in *Computational Logic in Mult-agent Systems (CLIMA'14)*, volume 8624 of *LNCS*, pp. 87–104. Springer, (2014).
[5] E. Black and A. Hunter, 'Reasons and options for updating an opponent model in persuasion dialogues', in *Theory and Applications of Formal Argumentation (TAFA'15)*, (2015).
[6] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet:, 'A comparative study of ranking-based semantics for abstract argumentation', in *Proceedings of AAAI'16*, pp. 914–920, (2016).
[7] M. Caminada and M. Podlaszewski, 'Grounded semantics as persuasion dialogue', in *Computational Models of Argument (COMMA'12)*, pp. 478–485, (2012).
[8] C. Cayrol and M. Lagasquie-Schiex, 'Graduality in argumentation', *Journal of Artificial Intelligence Research*, **23**, 245–297, (2005).
[9] F. Cerutti, N. Tintarev, and N. Oren, 'Formal arguments, preferences, and natural language interfaces to hhuman: An empirical evaluation', in *Proceedings of ECAI*, pp. 207–212, (2014).
[10] P. Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games', *Artificial Intelligence*, **77**, 321–357, (1995).
[11] P. Dung and P. Thang, 'Towards (probabilistic) argumentation for jury-based dispute resolution', in *Computational Models of Argument (COMMA'10)*, pp. 171–182. IOS Press, (2010).
[12] X. Fan and F. Toni, 'Assumption-based argumentation dialogues', in *Proceedings of IJCAI'11*, pp. 198–203, (2011).
[13] C. Hadjinikolis, Y. Siantos, S. Modgil, E. Black, and P. McBurney, 'Opponent modelling in persuasion dialogues', in *Proceedings of IJCAI*, pp. 164–170, (2013).
[14] E. Hadoux, A. Beynier, N. Maudet, P. Weng, and A. Hunter, 'Optimization of probabilistic argumentation with markov decision models', in *Proceedings of IJCAI'15*, (2015).
[15] A. Hunter, 'Some foundations for probabilistic argumentation', in *Computational Models of Argument (COMMA'12)*, pp. 117–128, (2012).
[16] A. Hunter, 'A probabilistic approach to modelling uncertain logical arguments', *International Journal of Approximate Reasoning*, **54**(1), 47–81, (2013).
[17] A. Hunter, 'Opportunities for argument-centric persuasion in behaviour change', in *Logics in Artificial Intelligence (JELIA'14)*, volume 8761 of *LNCS*, pp. 48–61. Springer, (2014).
[18] A. Hunter, 'Probabilistic strategies in dialogical argumentation', in *Scalable Uncertainty Management (SUM'14)*, volume 8720 of *LNCS*, pp. 190–202. Springer, (2014).
[19] A. Hunter, 'Modelling the persuadee in asymmetric argumentation dialogues for persuasion', in *Proceedings of IJCAI 2015*, (2015).
[20] A. Hunter and M. Thimm, 'Probabilistic argument graphs for argumentation lotteries', in *Computational Models of Argument*, pp. 313–324, (2014).
[21] A. Hunter and M. Thimm, 'Probabilistic argumentation with incomplete information', in *Proceedings of ECAI*, pp. 1033–1034, (2014).
[22] J. Leite and J. Martins, 'Social abstract argumentation', in *Proceedings of IJCAI'11*, (2011).

[23] H. Li, N. Oren, and T. Norman, 'Probabilistic argumentation frameworks', in *Theory and Applications of Formal Argumentation (TAFA'11)*, pp. 1–16, (2011).
[24] P. Matt and F. Toni, 'A game-theoretic measure of argument strength for abstract argumentation', in *Logics in Aartificial Intelligence (JELIA'08)*, volume 5293 of *LNCS*, pp. 285–297, (2008).
[25] N. Oren and T. Norman, 'Arguing using opponent models', in *Argumentation in Multi-agent Systems*, volume 6057 of *LNCS*, pp. 160–174, (2009).
[26] H. Prakken, 'Coherence and flexibility in dialogue games for argumentation', *Journal of Logic and Computation*, **15**(6), 1009–1040, (2005).
[27] H. Prakken, 'Formal sytems for persuasion dialogue', *Knowledge Engineering Review*, **21**(2), 163–188, (2006).
[28] A. Rago, F. Toni, M. Aurisicchio, and P. Baroni, 'Discontinuity-free decision support with quantitative argumentation debates', in *Proceedings of. KR'16*, pp. 63–73, (2016).
[29] I. Rahwan and K. Larson, 'Pareto optimality in abstract argumentation', in *Proceedings of AAAI 2008*, pp. 150–155. AAAI Press, (2008).
[30] I. Rahwan, M. Madakkatel, J. Bonnefon, R. Awan, and S. Abdallah, 'Behavioural experiments for assessing the abstract argumentation semantics of reinstatement', *Cognitive Science*, **34**(8), 14831502, (2010).
[31] T. Rienstra, M. Thimm, and N. Oren, 'Opponent models with uncertainty for strategic argumentation', in *Proceedings of IJCAI'13*, pp. 332–338. IJCAI/AAAI, (2013).
[32] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor, 'Heuristics in argumentation: A game theory investigation', in *Computational Models of Argument (COMMA 2008)*, pp. 324–335. IOS Press, (2008).
[33] A. Rosenfeld and S. Kraus, 'Providing arguments in discussions based on the prediction of human argumentative behavior', in *Proceedings of AAAI'15*, (2015).
[34] M. Thimm, 'A probabilistic semantics for abstract argumentation', in *Proceedings of ECAI'12*, pp. 750–755, (2012).
[35] M. Thimm, 'Strategic argumentation in multi-agent systems', *Kunstliche Intelligenz*, (2014).