# Updating Probabilistic Epistemic States in Persuasion Dialogues

Anthony Hunter[1] and Nico Potyka[2]

[1] Department of Computer Science, University College London, UK ,
[2] Institute of Cognitive Science, University of Osnabrück, Germany

**Abstract.** In persuasion dialogues, the ability of the persuader to model the persuadee allows the persuader to make better choices of move. The epistemic approach to probabilistic argumentation is a promising way of modelling the persuadee's belief in arguments, and proposals have been made for update methods that specify how these beliefs can be updated at each step of the dialogue. However, there is a need to better understand these proposals, and moreover, to gain insights into the space of possible update functions. So in this paper, we present a general framework for update functions in which we consider existing and novel update functions.

## 1 Introduction

The aim of persuasion is for the persuader to change the mind of the persuadee, and the provision of good arguments, and possibly counterarguments, is of central importance for this. Some recent developments in the field of computational persuasion have focused on the need to model the beliefs of the persuadee in order for the persuader to better select arguments to present to the persuadee. For instance, if the persuader wants to persuade the persuadee to give up smoking, and the persuader knows that the persuadee believes that if he gives up smoking, he will put on weight, then the persuader could start the dialogue by providing a counterargument to this, for example by saying that there is a local football team for ex-smokers who are looking for new players.

One approach to modelling the persuadee is to harness the epistemic approach to probabilitic argumentation [11]. In this, an argument graph (as definded by Dung [4]) is used to represent the arguments and attacks between them, and a probability distribution over the subsets of arguments is used to represent the uncertainty over which arguments are believed. The belief in an individual argument is then the sum of the belief in the subsets that contain this argument.

When a persuader starts a dialogue with a persuadee, the persuader identifies an appropriate probability distribution to represent what s/he thinks the persuadee believes. Then during the dialogue, the moves are made by the participants according to some protocol. After each move, the belief is updated using an update function (see Figure 1). Some initial proposals for update functions have been made (e.g. [10]) which seem intuitive and well-behaved, but there is a lack of a general understanding of what an update function is, of what the space of options are, and of how alternatives could be defined. The aim of this paper is to address these questions by proposing some basic

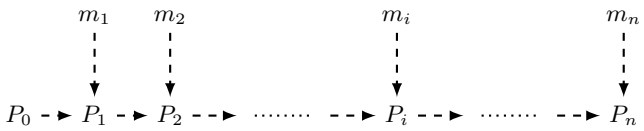


**Fig. 1.** Schematic representation of a dialogue $D = [m_1, \ldots, m_n]$ and user models $P_i$. Each user model $P_i$ is obtained from $P_{i-1}$ and move $m_i$ using an update method.

properties for update functions, and then proposing a framework for update measures in which we show how existing and some useful novel update functionss are situated.

## 2 Basics

We consider a finite argument graph $G$ with arguments $\mathrm{Args}$ and attacks $\mathrm{Attacks}$. For $A \in \mathrm{Args}$, we let $A^- = \{B \in \mathrm{Args} \mid (B, A) \in \mathrm{Attacks}\}$. $\mathrm{Form}$ denotes the set of propositional formulas over $\mathrm{Args}$. That is, $\mathrm{Form}$ is the smallest set that contains $\mathrm{Args}$ and is closed under application of the usual logical connectives like $\neg$ and $\wedge$. An interpretation of $\mathrm{Form}$ is a subset $X \subseteq \mathrm{Args}$. $X$ satisfies an atomic formula $A \in \mathrm{Args}$ iff $A \in X$ and we write $X \models A$ in this case. The satisfaction relation is extended to complex formulas in the usual way. For instance, $X \models F_1 \wedge F_2$ iff $X \models F_1$ and $X \models F_2$. A probability distribution over $\mathrm{Args}$ is a function $P : 2^{\mathrm{Args}} \to [0, 1]$ such that $\sum_{X \subseteq \mathrm{Args}} P(X) = 1$. We let $\mathcal{P}$ denote the set of all probability distributions over $\mathrm{Args}$. When speaking of topological properties of subsets of $\mathcal{P}$, we regard probability distributions as probability vectors and consider the usual topology on $\mathbb{R}^n$. Note that we can do so because $2^{\mathrm{Args}}$ is finite (because $\mathrm{Args}$ is finite). For $F \in \mathrm{Form}$, we let $P(F) = P(\{X \subseteq \mathrm{Args} \mid X \models F\})$. A complete conjunction over a subset $X \subseteq \mathrm{Args}$ is a conjunction of the form $\bigwedge_{A \in X} L_A$, where either $L_A = A$ or $L_A = \neg A$. Let $\mathrm{Conj}(X)$ denote the set of all complete conjunctions over $X$. In the following, we will make use of the fact that there is a 1-1 relationship between $\mathrm{Conj}(\mathrm{Args})$ and the interpretations $2^{\mathrm{Args}}$. More strictly speaking, a complete conjunction $\bigwedge_{A \in \mathrm{Args}} L_A$ corresponds to the interpretation $\{A \in \mathrm{Args} \mid L_A = A\}$ that contains all arguments that appear positive in the conjunction. Conversely, an interpretation $X \subseteq \mathrm{Args}$ corresponds to the complete conjunction $\bigwedge_{A \in X} A \wedge \bigwedge_{A \in \mathrm{Args} \setminus X} \neg A$.

Intuitively, a probability distribution over $\mathrm{Args}$ represents the epistemic state of an agent. Given an argument graph $G$, we want to impose certain constraints on probability distributions. We can consider some of the following rationality postulates for the epistemic state represented by $P$ [11].

- **RAT:** $P$ is *rational* iff for all $(A, B) \in \mathrm{Attacks}$, $P(A) > 0.5$ implies $P(B) \leq 0.5$.
- **COH:** $P$ is *coherent* iff for all $(A, B) \in \mathrm{Attacks}$, $P(A) \leq 1 - P(B)$.
- **SFOU:** $P$ is *semi-founded* iff $A^- = \emptyset$ implies $P(A) \geq 0.5$.
- **FOU:** $P$ is *founded* iff $A^- = \emptyset$ implies $P(A) = 1$.
- **SOPT:** $P$ is *semi-optimistic* iff $A^- \neq \emptyset$ implies $P(A) \geq 1 - \sum_{B \in A^-} P(B)$.
- **OPT:** $P$ is *optimistic* iff $P(A) \geq 1 - \sum_{B \in A^-} P(B)$.

For a subset $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT, JUS\}$ of rationality postulates, we write $P \models R$ iff $P$ satisfies all constraints in $R$ and for a subset $T \subseteq \mathcal{P}$, we write $T \models R$ iff $P \models R$ for all $P \in T$.

## 3 Properties of Update Functions

We can model the change of an agent's epistemic state in a dialogue by an update function [10]. Our goal here is to investigate the space of possible update functions systematically. Formally, we regard an update function as a function $U : \mathcal{P} \times \mathrm{Form} \to 2^{\mathcal{P}}$ that takes a probability distribution and a formula and maps them to a set of probability distributions $U(P, F)$ that satisfy $F$ in some way. In the following, we list several properties that might be interesting in this context. We start with a list of general properties.

- **Uniqueness:** $|U(P, F)| \leq 1$.
- **Completeness:** If $F \not\equiv \bot$ then $|U(P, F)| \geq 1$.
- **Tautology:** $U(P, \top) = \{P\}$.
- **Contradiction:** $U(P, \bot) = \emptyset$.
- **Representation Invariance:** If $F \equiv G$ then $U(P, F) = U(P, G)$.
- **Idempotence:** If $U(P, F) = \{P^*\}$ then $U(P^*, F) = \{P^*\}$.
- **Order Invariance:** $U(U(P, F_1), F_2) = U(U(P, F_2), F_1)$.

Uniqueness says that the solution of the update is always unique. Completeness says that a solution always exists when the new information is consistent. Tautology says that updating with a tautology should not change the epistemic state because we do not add any new information. Since our generated epistemic state should be consistent, Contradiction demands that updating with a contradictory formula should yield the empty set. Representation invariance says that semantically equivalent formulas should result in the same update. Idempotence says that if the update yields a unique solution, then updating again with the same information should not change the result. Order invariance says that the order in which we update does not affect the result.

Next, we consider some semantical properties. To begin with, we might want that updates take the structure of the argument graph into account. Therefore, we consider the following property for subsets $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT\}$ of rationality postulates:

- **R-Consistency:** If $P \models R$ then $U(P, F) \models R$.

In addition, the probability distributions in $U(P, F)$ should satisfy $F$ in some way. We consider the following satisfaction conditions.

- **STRICT:** $P$ satisfies $F$ *strictly* iff $P(F) = 1$.
- **$\epsilon$-WEAK:** $P$ satisfies $F$ *$\epsilon$-weakly* iff $P(F) \geq 0.5 + \epsilon$ for $\epsilon \in (0, 0.5)$.

*Remark 1.* Note that strict satisfaction implies $\epsilon$-weak satisfaction for all $\epsilon \in (0, 0.5)$.

For a satisfaction condition $S \in \{STRICT, \epsilon\text{-}WEAK\}$ and a formula $F \in \mathrm{Form}$, we write $P \models_S F$ iff $P$ satisfies $F$ with respect to $S$ and for a subset $T \subseteq \mathcal{P}$, we write $T \models_S F$ iff $P \models_S F$ for all $P \in T$. Analogous to rationality postulates, we consider the following property for $S \in \{STRICT, \epsilon\text{-}WEAK\}$:

– **S-Consistency:** $U(P, F) \models_S F$.

For a set of rationality postulates $R$ and a satisfaction condition $S$, we define the set of *R-S-models* of $F \in \mathrm{Form}$ by

$$\mathrm{Mod}_{R,S}(F) = \{P \in \mathcal{P} \mid P \models R, P \models_S F\}$$

We call $F$ *R-S-consistent* if $\mathrm{Mod}_{R,S}(F) \neq \emptyset$ and *R-S-inconsistent* otherwise. If $F$ is R-S-inconsistent, the condition of S-consistency becomes $\emptyset \models_S F$ and is trivially true. The following example illustrates an R-S-inconsistency.

*Example 1.* Consider an argument graph over $A, B$ with $\mathrm{Attacks} = \{(A, B)\}$. Let $R = \{RAT, FOU\}$. Then $FOU$ implies $P(A) = 1$ for all $P \in \mathrm{Mod}_{R,S}(\top)$ and therefore $RAT$ implies $P(B) \leq 0.5$. Hence, $\mathrm{Mod}_{R,\epsilon\text{-}WEAK}(B) = \emptyset$ for all $\epsilon > 0$.

Finally, we might want to update the epistemic state such that we minimally change the prior state. To this end, we can consider different change functions over $\mathcal{P}$. The first class of change measures that we consider measure the difference in probability mass that is assigned to interpretations.

– **Manhattan Distance:** $d_1(P, P^*) = \sum_{X \subseteq \mathrm{Args}} |P(X) - P^*(X)|$.
– **Least Squares Distance:** $d_2(P, P^*) = \sum_{X \subseteq \mathrm{Args}} (P(X) - P^*(X))^2$.
– **Maximum Distance:** $d_\infty(P, P^*) = \max_{X \subseteq \mathrm{Args}} |P(X) - P^*(X)|$.
– **KL-divergence:** $d_{KL}(P^*, P) = \sum_{X \subseteq \mathrm{Args}} P^*(X) \cdot \log \frac{P^*(X)}{P(X)}$.

Note that the KL-divergence is not a metric. In particular, it is asymmetric and we use the prior distribution $P$ as the second argument. If we have $P^*(X) > 0 = P(X)$ for some $X \subseteq \mathrm{Args}$, we let $d_{KL}(P^*, P) = \infty$ as usual.

When updating our belief with respect to a set of literals $\Phi$, we might be interested only in the change with respect to atoms not appearing in $\Phi$. The following two distance measures capture this intuition. Here, $X \subseteq \mathrm{Args}$ denotes a set of arguments that is supposed to be updated.

– **Atomic Distance:** $d^X_{\mathrm{At}}(P, P^*) = \sum_{B \in \mathrm{Args} \setminus X} |P(B) - P^*(B)|$.
– **Joint Distance:** $d^X_{\mathrm{Jo}}(P, P^*) = \sum_{C \in \mathrm{Conj}(\mathrm{Args} \setminus X)} |P(C) - P^*(C)|$.

Both measures can be zero even though the distributions are unequal. This happens, when they have equal marginal probabilities on $\mathrm{Args} \setminus X$ for the atomic distance measure and when they have equal marginal probabilities on $\mathrm{Conj}(\mathrm{Args} \setminus X)$ for the joint distance measure. Hence, they are not metrics. However, they are pseudometrics as we explain in the full version[3]. We illustrate the different change measures in Figure 1.

We consider the following minimality properties for each change measure $d$, set of rationality postulates $R$ and satisfaction condition $S$:

– **R-S-d-minimality:** If $P^* \in U(P, F)$, then $P^*$ minimizes the distance to $P$ over $\mathrm{Mod}_{R,S}(F)$.

R-S-d-minimality demands that we update in such a way that we minimize the distance to the prior distribution among all probability distributions that satisfy the argument graph and the new information with respect to the chosen semantics.

[3] Full version is at http://www0.cs.ucl.ac.uk/staff/a.hunter/papers/updatefunctionfull.pdf

| A | B | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $d$ | $d_1$ | $d_2$ | $d_\infty$ | $d_{KL}$ | $d_{At}^{\{A\}}$ | $d_{Jo}^{\{A\}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.3 | 0.4 | 0.5 | 0.5 | 0.3 | $d(P_0, P1)$ | 0.3 | 0.03 | 0.1 | 0.16 | 0.1 | 0.1 |
| 0 | 1 | 0.3 | 0.2 | 0.2 | 0.4 | 0.3 | $d(P_0, P2)$ | 0.4 | 0.06 | 0.2 | 0.09 | 0.1 | 0.2 |
| 1 | 0 | 0.3 | 0.2 | 0.2 | 0.1 | 0.2 | $d(P_0, P3)$ | 0.6 | 0.1 | 0.2 | $\infty$ | 0 | 0 |
| 1 | 1 | 0.1 | 0.1 | 0.1 | 0 | 0.2 | $d(P_0, P4)$ | 0.2 | 0.02 | 0.1 | 0.05 | 0.01 | 0.2 |

**Table 1.** Illustration of different change measures.

## 4 Refinement-Based Update functions

In [10], several update functions have been proposed that are defined by means of the following refinement function. They are restricted in the sense that they are defined only for literals.

**Definition 1.** *Let $L \in \mathsf{Formulae}(G)$ be a literal, let $P$ be a probability distribution, and let $\lambda \in [0,1]$. The* **refinement function** $H_\lambda : \mathcal{P} \times \{A, \neg A \mid A \in \mathrm{Args}\} \to \mathcal{P}$ *is defined by $H_\lambda(P, L) = P^*$ as follows where $X \subseteq \mathrm{Args}$*

$$P^*(X) = \begin{cases} P(X) + \lambda \cdot P(h_L(X)) & \text{if } X \models L \\ (1-\lambda) \cdot P(X) & \text{if } X \models \neg L, \end{cases}$$

*where $h_L(X) = X \setminus \{A\}$ if $L = A$ and $h_L(X) = X \cup \{A\}$ if $L = \neg A$ for some $A \in \mathrm{Args}$.*

If we think of interpretations as bit vectors $(b_1, \ldots, b_n)$ where $b_i$ is the truth state of the $i$-th argument, redistribution with respect to $A_i$ can be explained as follows: for each bit vector $(b_1, \ldots, b_n)$, if $b_i = 1$, then move a fraction $\lambda$ of the probability mass of $(b_1, \ldots, b_{i-1}, 0, b_{i+1}, \ldots, b_n)$ to $(b_1, \ldots, b_n)$. We illustrate this in Table 2.

Let us note that refinement functions are actually commutative in the sense that $H_{\lambda_2}(H_{\lambda_1}(P, L_1), L_2) = H_{\lambda_1}(H_{\lambda_2}(P, L_2), L_1)$, see [10], Proposition 8. Since the order in which we add literals is not important, refinement functions can also be applied to sets of literals $\Phi$ recursively, where we let $H_\lambda(P, \emptyset) = P$ and $H_\lambda(P, \Phi \cup \{L\}) = H_\lambda(H_\lambda(P, L), \Phi)$. As the following lemma explains, for $\lambda = 1$, updating with multiple literals comes down to shifting probability mass to the interpretations that satisfy the conjunction of these literals.

**Lemma 1.** *Let $X = \{A_1, \ldots, A_k\} \subseteq \mathrm{Args}$ and for $i = 1, \ldots, k$, let $L_i \in \{A_i, \neg A_i\}$. Let $P$ be a probability distribution and let $H_1(P, \{L_1, \ldots, L_k\}) = P^*$. Then for all $C \in \mathrm{Conj}(X)$ and $D \in \mathrm{Conj}(\mathrm{Args} \setminus X)$,*

$$P^*(C \wedge D) = \begin{cases} P(C \wedge D) + \sum_{C' \in \mathrm{Conj}(X) \setminus \{C\}} P(C' \wedge D) & \text{if } C = \bigwedge_{i=1}^k L_i \\ 0 & \text{else.} \end{cases}$$

We will now analyze some refinement-based update functions from [10] by means of the properties introduced in the previous section. Since the refinement-based update

| A | B | C | $P$ | $H_{0.75}(P,A)$ | $H_1(P,A)$ | $U_{\mathrm{na}}(P,B)$ | $U_{\mathrm{tr}}(P,B)$ | $U_{\mathrm{tr}}(P,A)$ | $U_{\mathrm{st}}(P,B)$ | $U_{\mathrm{st}}(P,A)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0.2 | 0.05 | 0 | 0 | 0 | 0 | 0 | 0.2 |
| 0 | 1 | 0 | 0.5 | 0.125 | 0 | 0.7 | 1 | 0 | 0.7 | 0.5 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0.1 | 0.025 | 0 | 0.1 | 0 | 0 | 0.3 | 0.1 |
| 1 | 0 | 0 | 0 | 0.15 | 0.2 | 0 | 0 | 0.7 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0.375 | 0.5 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0.1 | 0.1 | 0.1 | 0 | 0 | 0.3 | 0 | 0.1 |
| 1 | 1 | 1 | 0.1 | 0.175 | 0.2 | 0.2 | 0 | 0 | 0 | 0.1 |

**Table 2.** Illustration of refinement-based updates for a graph with $C$ attacks $B$ and $B$ attacks $A$. Note, by definition, $H_1(P,A) = U_{\mathrm{na}}(P,A)$ and $H_1(P,B) = U_{\mathrm{na}}(P,B)$.

functions are only defined for atoms or literals, Tautology, Contradiction and Representation Invariance are not interesting here. However, it is reasonable to consider Idempotence and Order Invariance restricted to literals.

The naive update function shifts the probability mass from an interpretation $X$ that violates $L$ to the corresponding interpretation that is obtained from $X$ by flipping the truth state of the argument in $L$.

**Definition 2 ([10]).** *The* **naive update function** $U_{\mathrm{na}} : \mathcal{P} \times \{A, \neg A \mid A \in \mathrm{Args}\} \to \mathcal{P}$ *is defined by* $U_{\mathrm{na}}(P, L) = H_1(P, L)$.

$U_{\mathrm{na}}$ satisfies the following properties.

**Proposition 1.** *$U_{\mathrm{na}}$ satisfies Uniqueness, Completeness, Idempotence, Order Invariance and STRICT-Satisfaction.*

The naive update function is intended to model persuadees who believe any arguments that are posited in a dialogue. The function does not take the structure of the argument graph into account, and therefore can generally violate all rationality postulates that we introduced over argument graphs. However, given an update literal over the argument $A$, the naive update is guaranteed to be minimal with respect to $d_{Jo}^{\{A\}}$ - in fact, the change with respect to $d_{Jo}^{\{A\}}$ is 0 as we show in the full version of this paper.

The next two update functions maintain consistency with the argument graph by also considering arguments that are connected to the argument whose state we update. They are restricted to atomic arguments, however.

The trusting update refines the naive update by also shifting the probability mass from all interpretations that satisfy the attackers and attackees of the update argument.

**Definition 3 ([10]).** *The* **trusting update function** $U_{\mathrm{tr}} : \mathcal{P} \times \mathrm{Args} \to \mathcal{P}$ *is defined by* $U_{\mathrm{tr}}(P, A) = H_1(P, \Phi)$*, where* $\Phi = \{A\} \cup \{\neg C \mid (A, C) \in \mathsf{Attacks}(G)$ *or* $(C, A) \in \mathsf{Attacks}(G)\}$.

$U_{\mathrm{tr}}$ satisfies the following properties.

**Proposition 2.** *$U_{\mathrm{tr}}$ satisfies Uniqueness, Completeness, Idempotence, Order Invariance, STRICT-Satisfaction and R-Satisfaction for all $R \subseteq \{RAT, COH\}$.*

$U_{\mathrm{tr}}$ can violate the remaining R-Satisfaction properties, but it does guarantee that the joint distance to the prior distribution is $0$. However, the joint distance is now not only defined with respect to the update argument, but also with all of its attackers and attackees as we show in the full version.

The strict update function conditionally updates the probability of an argument to $1$. In order to maintain consistency with the argument graph, the update is only performed if no attackers of the argument are believed in the current epistemic state. If the update is performed, the belief in attacked arguments will additionally be set to $0$.

**Definition 4 ([10]).** *The* **strict update function** *is a function* $U_{\mathrm{st}} : \mathcal{P} \times \mathrm{Args} \to \mathcal{P}$. *For* $A \in \mathrm{Args}$*, let* $\Phi = \{A\} \cup \{\neg C \mid (A, C) \in \mathrm{Attacks}\}$ *and let the constraint* $C(P)$ *be true iff for all* $(B, A) \in \mathrm{Attacks}$*,* $P(B) \leq 0.5$*. Then* $U_{\mathrm{st}}(P, A) = P^*$ *where*

$$P^* = \begin{cases} H_1(P, \Phi) & \text{if } C(P) \\ P & \text{else} \end{cases}$$

$U_{\mathrm{st}}$ satisfies the following properties.

**Proposition 3.** $U_{\mathrm{st}}$ *satisfies Uniqueness, Completeness, Idempotence and R-Satisfaction for all* $R \subseteq \{RAT, COH, SFOU, FOU\}$.

$U_{\mathrm{st}}$ does not satisfy Order Invariance, but it satisfies all semantical constraints except R-OPT and R-SOPT. $U_{\mathrm{st}}$ again guarantees joint distance $0$, this time with respect to the update argument and all of its attackees. We refer again to the full version of this paper for more details and proofs.

In [10], $H_{0.75}$ is considered as an alternative to $H_1$ in the above definition, and this is used to model skeptical agents who do not entirely believe an argument when updating.

## 5 R-S-d Update Functions

We now consider another class of update functions. Whereas refinement-based update functions are based on the idea of shifting probability mass in a specific way, we will now consider a more declarative approach using tools from numerical optimization. R-S-d Update Functions are defined by minimizing some notion of distance subject to semantical constraints.

**Definition 5.** *Let* $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT\}$*,* $S \in \{STRICT, \epsilon\text{-}WEAK\}$ *and* $d \in \{d_1, d_2, d_\infty, d^X_{\mathrm{At}}, d^X_{\mathrm{Jo}}\}$*. An* **R-S-d Update Function** $U_{R,S,d} : \mathcal{P} \times \mathrm{Form} \to 2^{\mathcal{P}}$ *is defined by*

$$U_{R,S,d}(P, F) = \arg \min_{P' \in \mathrm{Mod}_{R,S}(F)} d(P, P').$$

Let us first note that most R-S-d update functions have some nice analytical properties.

| A | B | C | $P$ | $U_{R_1,S,d}(P,A)$ | $U_{R_1,S,d}(P,B)$ | $U_{R_2,S,d}(P,A)$ | $U_{R_2,S,d}(P,\top)$ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0.2 | 0 | 0 | 0 | 0.17 |
| 0 | 1 | 0 | 0.5 | 0 | 1 | 0 | 0.49 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0.1 | 0 | 0 | 0 | 0.07 |
| 1 | 0 | 0 | 0 | 0.45 | 0 | 0 | 0.02 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0.5 |
| 1 | 0 | 1 | 0.1 | 0.55 | 0 | 1 | 0.09 |
| 1 | 1 | 1 | 0.1 | 0 | 0 | 0 | 0.12 |

**Table 3.** Illustration of R-S-d updates with $R_1 = \{COH\}$, $R_2 = \{COH, SOPT\}$, $S = STRICT$ and $d = d_2$.

**Lemma 2.** *For each $R \subseteq \{COH, SFOU, FOU, SOPT, OPT\}$ (we left out RAT), $S \in \{STRICT, \epsilon\text{-WEAK}\}$ and $d \in \{d_1, d_2, d_\infty, d_m, d_{KL}, d^X_{\mathrm{At}}, d^X_{\mathrm{Jo}}\}$, computing $U_{R,S,d}(P,F)$ corresponds to a convex combination problem. In particular, the set $U_{R,S,d}(P,F)$ will be non-empty, convex and compact whenever $\mathrm{Mod}_{R,S}(F) \neq \emptyset$.*

*If $R$ includes RAT, $U_{R,S,d}(P,F)$ will be non-empty and compact whenever $\mathrm{Mod}_{R,S}(F) \neq \emptyset$.*

We have the following general guarantees for R-S-d update functions.

**Proposition 4.** *For all $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT\}$, $S \in \{STRICT, \epsilon\text{-WEAK}\}$ and $d \in \{d_1, d_2, d_\infty, d_m, d^X_{\mathrm{At}}, d^X_{\mathrm{Jo}}\}$, $U_{R,S,d}$ satisfies Completeness (if the update argument is R-S-consistent), R-consistency, S-consistency and R-S-d-minimality.*

*If we exclude RAT from R and $d \in \{d_2, d_{KL}\}$, $U_{R,S,d}$ also satisfies Uniqueness, Tautology, Contradiction, Representation Invariance and Idempotence.*

We can give some stronger guarantees for some special cases, see the full paper for a detailed analysis.

Order Invariance can be violated for many combinations of semantical constraints and change measures. We give a simple example for the Euclidean distance without semantical constraints on the argument graph.

*Example 2.* Consider an argument graph over $A, B$, let $R = \emptyset$, $S = STRICT$ and $d = d_2$. Let $P$ be defined by $P(\{B\}) = 0.5$, $P(\{A,B\}) = 0.5$. Then $P_1 = U_{R,S,d}(U_{R,S,d}(P,A),B)$ is given by $P_1(\{B\}) = 0.125$, $P_1(\{A,B\}) = 0.875$, whereas $P_2 = U_{R,S,d}(U_{R,S,d}(P,B),A)$ is given by $P_2(\{A\}) = 0.25$, $P_2(\{A,B\}) = 0.75$.

What can we say about the relationship between refinement-based update functions and R-S-d update functions? We first note that $R$-$S$-$d$-update functions generalize the naive update function in the following sense.

**Proposition 5.** *Consider an arbitrary set of semantical constraints $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT\}$, a probability distribution $P \in \mathcal{P}$ and let $L \in \{A, \neg A\}$ be a literal for some $A \in \mathrm{Args}$. If there is a $P^* \in \mathrm{Mod}_{R,STRICT}(L)$ such that $d^{\{A\}}_{Jo}(P,P^*) = 0$ then $U_{R,STRICT,d^{\{A\}}_{Jo}}(P,L) = \{U_{\mathrm{na}}(P,L)\}$.*

*Remark 2.* Note that if there is no $P^* \in \mathrm{Mod}_{R,STRICT}(L)$ such that $d_{Jo}^{\{A\}}(P, P^*) = 0$, then applying the Naive update function will violate some semantical constraint in $R$ (because the probability distribution resulting from the naive update will have distance 0). Hence, $U_{R,STRICT,d_{Jo}^{\{A\}}}$ agrees with $U_{\mathrm{na}}$ whenever $U_{\mathrm{na}}$ is consistent with $R$. Otherwise, $U_{R,STRICT,d_{Jo}^{\{A\}}}$ will select probability distributions that are consistent with $R$ and minimize the joint distance.

In particular, the Naive update function can be thought of as a special case of the following $R$-$S$-$d$-update function.

**Corollary 1.** $U_{\emptyset,STRICT,d_{Jo}^{\{A\}}}(P, L) = \{U_{\mathrm{na}}(P, L)\}$.

The trusting method can similarly be generalized by an $R$-$S$-$d$-update function.

**Proposition 6.** *Consider an arbitrary set of semantical constraints* $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT\}$*, a probability distribution* $P \in \mathcal{P}$ *and let* $L \in \{A, \neg A\}$ *be a literal for some* $A \in \mathrm{Args}$*. Let* $X' = \{C \mid (A, C) \in \mathsf{Attacks}(G)$ *or* $(C, A) \in \mathsf{Attacks}(G)\}$ *and* $X = \{A\} \cup X'$*. If there is a* $P^* \in \mathrm{Mod}_{R,STRICT}(L)$ *such that* $d_{Jo}^{X}(P, P^*) = 0$ *then* $U_{R,STRICT,d_{Jo}^{X}}(P, L \wedge \bigwedge_{C \in X'} \neg C) = \{U_{\mathrm{tr}}(P, L)\}$.

**Corollary 2.** $U_{\emptyset,STRICT,d_{Jo}^{X}}(P, L \wedge \bigwedge_{C \in X'} \neg C) = \{U_{\mathrm{tr}}(P, L)\}$.

We could get a similar result for the strict update using the joint distance over the update argument and its attackees. This would require a case differentiation analogous to the case differentation that is used for the strict update.

## 6 Conclusions and Future Work

Most proposals for dialogical argumentation focus on protocols (*e.g.*, [14], [15], [5], [2]) with strategies being under-developed. See [18] for a review of strategies in multi-agent argumentation. There are proposals for modelling the likelihood of the moves that an opposing agent might make (e.g. [16, 6, 7, 17]). Note, however, that none of the above proposals consider the beliefs of the opposing agent. In [1], a planning system is used by the persuader to optimize choice of arguments based on belief in premises. However, there is no consideration of how the beliefs are updated during the dialogue.

The epistemic approach to probabilistic argumentation offers a formal framework for modelling a persuadee's beliefs in arguments. There are methods for updating beliefs during a dialogue [10], for efficient representation and reasoning with the pesuadee model [8], and for harnessing decision-theoretic decision rules for optimizing the choice of arguments based on the persuadee model [9]. Therefore, the framework for update functions presented in this paper clarifies and extends the space of update functions that we can harness in persuasion dialogues.

There are several interesting directions for future work. First, we can investigate different ways to deal with the problem of non-unique solutions. We might focus on some best solution or represent epistemic states by sets of probability distributions rather than by a single one. Second, we can deal with inconsistencies like in Example 1 in different ways. We might consider priorities over different semantical constraints [12] or select

solutions that violate the constraints in a minimal way [3, 13]. Third, we can try to include more expressive argumentation frameworks by introducing numerical constraints for other relations than attack relations.

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

## References

1. Black, E., Coles, A., Bernardini, S.: Automated planning of simple persuasion dialogues. In: Proc. of CLIMA'14. LNCS, vol. 8624, pp. 87–104. Springer (2014)
2. Caminada, M., Podlaszewski, M.: Grounded semantics as persuasion dialogue. In: Proc. Int. Conf. Computational Models of Argument (COMMA). pp. 478–485 (2012)
3. Daniel, L.: Paraconsistent Probabilistic Reasoning. Ph.D. thesis, L'École Nationale Supérieure des Mines de Paris (2009)
4. Dung, P.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. Artificial Intelligence 77, 321–357 (1995)
5. Fan, X., Toni, F.: Assumption-based argumentation dialogues. In: Proc IJCAI'11. pp. 198–203 (2011)
6. Hadjinikolis, C., Siantos, Y., Modgil, S., Black, E., McBurney, P.: Opponent modelling in persuasion dialogues. In: Proc. IJCAI'13. pp. 164–170 (2013)
7. Hadoux, E., Beynier, A., Maudet, N., Weng, P., Hunter, A.: Optimization of probabilistic argumentation with Markov decision models. In: Proc IJCAI'15. pp. 2004–2010 (2015)
8. Hadoux, E., Hunter, A.: Computationally viable handling of beliefs in arguments for persuasion. In: Proc. of ICTAI'16. pp. 319–326. IEEE Press (2016)
9. Hadoux, E., Hunter, A.: Strategic sequences of arguments for persuasion using decision trees. In: Proc. of AAAI'17. AAAI Press (2017), in press
10. Hunter, A.: Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In: Proc. IJCAI'15. pp. 3055–3061 (2015)
11. Hunter, A., Thimm, M.: On partial information and contradictions in probabilistic abstract argumentation. In: Principles of Knowledge Representation and Reasoning (KR'16). pp. 53–62 (2016)
12. Potyka, N.: Reasoning over linear probabilistic knowledge bases with priorities. In: International Conference on Scalable Uncertainty Management. LNCS, vol. 9310, pp. 121–136. Springer (2015)
13. Potyka, N., Thimm, M.: Probabilistic reasoning with inconsistent beliefs using inconsistency measures. In: Proc. IJCAI'15. pp. 3156–3163 (2015)
14. Prakken, H.: Coherence and flexibility in dialogue games for argumentation. Journal of Logic and Computation 15(6), 1009–1040 (2005)
15. Prakken, H.: Formal sytems for persuasion dialogue. Knowledge Engineering Review 21(2), 163–188 (2006)
16. Rienstra, T., Thimm, M., Oren, N.: Opponent models with uncertainty for strategic argumentation. In: Proc. IJCAI'13. pp. 332–338 (2013)
17. Rosenfeld, A., Kraus., S.: Providing arguments in discussions on the basis of the prediction of human argumentative behavior. ACM Transactions on Interactive Intelligent Systems 6, 30:1–30:33 (2016)
18. Thimm, M.: Strategic argumentation in multi-agent systems. Kunstliche Intelligenz 28, 159–168 (2014)