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 probabilistic argumentation [11]. In this, an argument graph (as defined 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
epistemic state represented by probability distributions. We can consider some of the following rationality postulates for the conjunction. Conversely, an interpretation of the fact that there is a 1-1 relationship between $X \cong \bigwedge A \in X$ and is closed under application of the usual logical connectives like $\neg$ and $\land$. An interpretation of $\text{Form}$ is a subset $X \subseteq \text{Args}$. $X$ satisfies an atomic formula $A \in \text{Args}$ 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, $X \models F_1 \land F_2$ iff $X \models F_1$ and $X \models F_2$. A probability distribution over $\text{Args}$ is a function $P : 2^\text{Args} \to [0,1]$ such that $\sum_{X \subseteq \text{Args}} P(X) = 1$. We let $\mathcal{P}$ denote the set of all probability distributions over $\text{Args}$. 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 $2^\text{Args}$ is finite (because $\text{Args}$ is finite). For $F \in \text{Form}$, we let $P(F) = P\{X \subseteq \text{Args} \mid X \models F\}$. A complete conjunction over a subset $X \subseteq \text{Args}$ is a conjunction of the form $\land_{A \in X} L_A$, where either $L_A = A$ or $L_A = \neg A$. Let $\text{Conj}(X)$ 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 $\text{Conj}(\text{Args})$ and the interpretations $2^\text{Args}$. More strictly speaking, a complete conjunction $\land_{A \in \text{Args}} L_A$ corresponds to the interpretation $\{A \in \text{Args} \mid L_A = A\}$ that contains all arguments that appear positive in the conjunction. Conversely, an interpretation $X \subseteq \text{Args}$ corresponds to the complete conjunction $\land_{A \in X} L_A \land \land_{A \in \text{Args}\setminus X} \neg A$. Intuitively, a probability distribution over $\text{Args}$ 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 $(A,B) \in \text{Attacks}$, $P(A) > 0.5$ implies $P(B) \leq 0.5$.
- **COH:** $P$ is **coherent** iff for all $(A,B) \in \text{Attacks}$, $P(A) \leq 1 - P(B)$.
- **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 $P(A) \geq 1 - \sum_{B \in A^-} P(B)$.

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.

2 Basics

We consider a finite argument graph $G$ with arguments $\text{Args}$ and attacks $\text{Attacks}$. For $A \in \text{Args}$, we let $A^- = \{B \in \text{Args} \mid (B,A) \in \text{Attacks}\}$. Form denotes the set of propositional formulas over $\text{Args}$. That is, $\text{Form}$ is the smallest set that contains $\text{Args}$ and is closed under application of the usual logical connectives like $\neg$ and $\land$. An interpretation of $\text{Form}$ is a subset $X \subseteq \text{Args}$. $X$ satisfies an atomic formula $A \in \text{Args}$ 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, $X \models F_1 \land F_2$ iff $X \models F_1$ and $X \models F_2$. A probability distribution over $\text{Args}$ is a function $P : 2^\text{Args} \to [0,1]$ such that $\sum_{X \subseteq \text{Args}} P(X) = 1$. We let $\mathcal{P}$ denote the set of all probability distributions over $\text{Args}$. 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 $2^\text{Args}$ is finite (because $\text{Args}$ is finite). For $F \in \text{Form}$, we let $P(F) = P\{X \subseteq \text{Args} \mid X \models F\}$. A complete conjunction over a subset $X \subseteq \text{Args}$ is a conjunction of the form $\land_{A \in X} L_A$, where either $L_A = A$ or $L_A = \neg A$. Let $\text{Conj}(X)$ 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 $\text{Conj}(\text{Args})$ and the interpretations $2^\text{Args}$. More strictly speaking, a complete conjunction $\land_{A \in \text{Args}} L_A$ corresponds to the interpretation $\{A \in \text{Args} \mid L_A = A\}$ that contains all arguments that appear positive in the conjunction. Conversely, an interpretation $X \subseteq \text{Args}$ corresponds to the complete conjunction $\land_{A \in X} L_A \land \land_{A \in \text{Args}\setminus X} \neg A$. Intuitively, a probability distribution over $\text{Args}$ 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 $(A,B) \in \text{Attacks}$, $P(A) > 0.5$ implies $P(B) \leq 0.5$.
- **COH:** $P$ is **coherent** iff for all $(A,B) \in \text{Attacks}$, $P(A) \leq 1 - P(B)$.
- **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 $P(A) \geq 1 - \sum_{B \in A^-} P(B)$. 

![Schematic representation of a dialogue](image)
For a subset $R \subseteq \{\text{RAT, COH, SFOU, FOU, SOPT, OPT, JUS}\}$ 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 $U: \mathcal{P} \times \text{Form} \rightarrow 2^\mathcal{P}$ 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 $R \subseteq \{\text{RAT, COH, SFOU, FOU, SOPT, OPT}\}$ 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 $P(F) = 1$.
- **ε-WEAK:** $P$ satisfies $F$ ε-weakly iff $P(F) \geq 0.5 + \epsilon$ for $\epsilon \in (0, 0.5)$.

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

For a satisfaction condition $S \in \{\text{STRICT, ε-WEAK}\}$ and a formula $F \in \text{Form}$, 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 $S \in \{\text{STRICT, ε-WEAK}\}$:
\(-\textbf{S-Consistency:} \ U(P, F) \models_S F.\)

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

\[
\text{Mod}_{R,S}(F) = \{ P \in \mathcal{P} \mid P \models_R P \models_S F \}
\]

We call \( F \) \( R-S\)-consistent if \( \text{Mod}_{R,S}(F) \neq \emptyset \) 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.

\textbf{Example 1.} Consider an argument graph over \( A, B \) with Attacks = \{\( (A, B) \)\}. Let \( R = \{\text{RAT}, \text{FOU}\} \). Then \( \text{FOU} \) implies \( P(A) = 1 \) for all \( P \in \text{Mod}_{R,S}(\top) \) and therefore \( \text{RAT} \) implies \( P(B) \leq 0.5 \). Hence, \( \text{Mod}_{R,\epsilonWEAK}(B) = \emptyset \) 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 \( P \). The first class of change measures that we consider measure the difference in probability mass that is assigned to interpretations.

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

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 \( X \subseteq \text{Args} \), we let \( d_{KL}(P^*, P) = \infty \) as usual.

Each distance measure focuses on different aspects of the redistribution of probability mass. In order to illustrate this, consider the probability vector \((0.2, 0.2, 0.2, 0.4)\) that assigns probabilities to four interpretations. Both the vectors \((0.3, 0.3, 0.3, 0.1)\) and \((0.25, 0.25, 0.25, 0.25)\) have Manhattan distance 0.4 to the original distribution because the same total amount of probability mass is shifted. However, if we want that the shift is equally distributed among the interpretations, we should use the least squares distance that penalizes higher deviations more heavily due to the quadratic exponent. Indeed, the least squares distance for the first vector is 0.12, whereas it is only 0.03 for the second one. If, on the other hand, we were only interested in the maximum absolute shift, we should apply the maximum distance. For instance, \((0.3, 0.3, 0.3, 0.1)\) has maximum distance 0.3 because of the shift in the fourth world. \((0.4, 0.5, 0.1)\) has also maximum distance 0.3 from \((0.2, 0.2, 0.2, 0.4)\) because the probability mass for no interpretation is shifted by more than 0.3. The KL-divergence measures the weighted difference in information content. Formally, the information content of an event \( \omega \) can be measured by the negative logarithm of its probability \( -\log P(\omega) \), see [12] for instance. This definition has some intuitive properties. For instance, as the event’s probability goes to 0, its information content goes to infinity (less likely means more surprising) and as its probability goes to 1, its information content goes to 0 (more likely means less surprising). The KL-divergence sums up terms.
\( P^*(X) \cdot \log \frac{P^*(X)}{P(X)} = P^*(X) \cdot (\log P(X) - (\log P^*(X)) \) That is, for each interpretation \( X \), it measures the difference in information content and weighs the difference by the probability of \( X \) with respect to \( P^* \).

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, \( X \subseteq \text{Args} \) denotes a set of arguments that is supposed to be updated.

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

Note that for \( \Phi = \emptyset \), \( d^X_{\text{At}} = d_1 \) because of the relationship between complete conjunctions over \( \text{Args} \) and interpretations that we explained before. However, in general, neither the atomic distance measure nor the joint distance measure are metrics. In particular, the distance can be zero, even though the distributions are unequal. This happens, when they have equal marginal probabilities on \( \text{Args} \setminus \Phi \) for the atomic distance measure and when they have equal marginal probabilities on \( \text{Conj}(\text{Args} \setminus \Phi) \) for the joint distance measure. However, both change measures are pseudometrics, that is, non-negative, symmetric functions that satisfy the triangle inequality. Just like our other change measures, they are continuous and convex in one argument when fixing the other.

**Lemma 1.** \( d^X_{\text{At}} \) and \( d^X_{\text{Jo}} \) are continuous and convex pseudometrics.

**Proof.** Non-negativity and Symmetry follow immediately from the definition. The triangle inequality follows from observing that \( |P_1(F) - P_2(F)| \leq |P_1(F) - P(F)| + |P(F) - P_2(F)| \) for all \( F \in \text{Form} \). Putting this into the definition, we get \( d^X_{\text{At}}(P_1, P_2) \leq d^X_{\text{At}}(P_1, P) + d^X_{\text{At}}(P, P_2) \) and \( d^X_{\text{Jo}}(P_1, P_2) \leq d^X_{\text{Jo}}(P_1, P) + d^X_{\text{Jo}}(P, P_2) \).

Continuity follows from the fact that both measures are composed of continuous functions of the arguments. Convexity of \( d^X_{\text{At}} \) for the first argument follows from observing that

\[
\begin{align*}
 d^X_{\text{At}}(\lambda P_1 + (1 - \lambda)P_2, P^*) &= \sum_{B \in \text{Args} \setminus \Phi} |(\lambda P_1(B) + (1 - \lambda)P_2(B)) - P^*(B)| \\
 &= \sum_{B \in \text{Args} \setminus \Phi} |\lambda(P_1(B) - P^*(B))| + (1 - \lambda)(P_2(B) - P^*(B))| \\
 &\leq \lambda \sum_{B \in \text{Args} \setminus \Phi} |(P_1(B) - P^*(B))| + (1 - \lambda) \sum_{B \in \text{Args} \setminus \Phi} |(P_2(B) - P^*(B))| \\
 &= \lambda \cdot d^X_{\text{At}}(P_1, P^*) + (1 - \lambda) \cdot d^X_{\text{At}}(P_2, P^*).
\end{align*}
\]

The argumentation is analogous for the second argument and \( d^X_{\text{Jo}} \). \( \square \)

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 \( \text{Mod}_{R,S}(F) \). 

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

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

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

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 2. Let $X = \{A_1, \ldots, A_k\} \subseteq \text{Args}$ 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 $C \in \text{Conj}(X)$ and $D \in \text{Conj}(\text{Args} \setminus X)$,

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

Table 1. Illustration of different change measures.
Proof. We prove the claim by induction over \( k \). We will restrict to the case \( L_k = A_k \). The case \( L_k = \neg A_k \) is analogous. For \( X = \{A_1\} \), we get from the definition of \( H_1 \) and the connection between interpretations and complete conjunctions over \( \text{Args} \) that for all \( D \in \text{Conj}(\text{Args} \setminus \{A_1\}) \), \( P^*(A_1 \land D) = P(A_1 \land D) + P(\neg A_1 \land D) = P(A_1 \land D) + \sum_{C' \in \{A_1, \neg A_1\} \setminus \{A_1\}} P(C' \land D) \) and \( P^*(\neg A_1 \land D) = 0 \).

For the induction step, consider \( X = \{L_1, \ldots, L_k, A_{k+1}\} \), where \( L_i \) contains argument \( A_i \) and let \( P_k = H_1(P, \{L_1, \ldots, L_k\}) \). By induction assumption, we have that for all \( C \in \text{Conj}(\text{Args} \setminus \{A_1, \ldots, A_k\}) \) and \( D \in \text{Conj}(\text{Args} \setminus \{A_1, \ldots, A_k\}) \),

\[
P_k(C \land D) = \begin{cases} P(C \land D) + \sum_{C' \in \text{Conj}(\{A_1, \ldots, A_k\}) \setminus \{C\}} P(C' \land D) & \text{if } C = \bigwedge_{i=1}^k L_i \\ 0 & \text{else.} \end{cases}
\]

We have \( H_1(P, X) = H_1(H_1(P, \{L_1, \ldots, L_k\}, \{A_{k+1}\})) = H_1(P_k, \{A_{k+1}\}) \) by commutativity of \( H_1 \) as discussed before. Let \( P^* = H_1(P, X) \) and let us check that the claim is true for \( P^* \) by looking at all possible complete conjunctions.

From the definition of \( H_1 \), we have for all \( C \in \text{Conj}(X \setminus \{A_{k+1}\}) \) and \( D \in \text{Conj}(\text{Args} \setminus X) \) that \( P^*(A_{k+1} \lor C \land D) = P_k(A_{k+1} \lor C \land D) + P_k(\neg A_{k+1} \land C \land D) \) and \( P^*(\neg A_{k+1} \land C \land D) = 0 \). We also know from our induction assumption that, if \( C \neq \bigwedge_{i=1}^k L_i \), then \( P_k(A_{k+1} \land C \land D) = 0 \) and \( P_k(\neg A_{k+1} \land C \land D) = 0 \) and therefore \( P^*(A_{k+1} \land C \land D) = 0 \). Hence, whenever \( C \neq \bigwedge_{i=1}^k L_i \land A_{k+1} \), then \( P^*(C \land D) = 0 \).

The only remaining case is

\[
P^*(\bigwedge_{i=1}^k L_i \land A_{k+1} \land D)
= P_k(\bigwedge_{i=1}^k L_i \land A_{k+1} \land D) + P_k(\bigwedge_{i=1}^k L_i \land \neg A_{k+1} \land D).
\]

Let us rewrite the two terms in (1). For the first term in (1), we have

\[
P_k(\bigwedge_{i=1}^k L_i \land A_{k+1} \land D)
= P(\bigwedge_{i=1}^k L_i \land A_{k+1} \land D) + \sum_{C' \in \text{Conj}(\{A_1, \ldots, A_k\}) \setminus \{A_{k+1}\}} P(C' \land A_{k+1} \land D)
\]

by our induction assumption. Similarly, for the second term in (1), we have

\[
P_k(\bigwedge_{i=1}^k L_i \land \neg A_{k+1} \land D)
= P(\bigwedge_{i=1}^k L_i \land \neg A_{k+1} \land D) + \sum_{C' \in \text{Conj}(\{A_1, \ldots, A_k\}) \setminus \{A_{k+1}\}} P(C' \land \neg A_{k+1} \land D).
\]
Table 2. Illustration of refinement-based updates for a graph with $C$ attacks $B$ and $B$ attacks $A$. Note, by definition, $H_1(P, A) = U_{na}(P, A)$ and $H_1(P, B) = U_{na}(P, B)$.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>$H_{0.75}(P, A)$</th>
<th>$H_1(P, A)$</th>
<th>$U_{na}(P, B)$</th>
<th>$U_{tr}(P, B)$</th>
<th>$U_{st}(P, A)$</th>
<th>$U_{st}(P, B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.125</td>
<td>0</td>
<td>0.7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>0.025</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.15</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.375</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>0.175</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Adding the second term from (2) to (3) gives us

$$\sum_{C' \in \text{Conj}(\{A_1, \ldots, A_k\}) \setminus \{A_{k+1} \}} P(C' \land A_{k+1} \land D)$$

$$+ P(\bigwedge_{i=1}^k L_i \land \neg A_{k+1} \land D) + \sum_{C' \in \text{Conj}(\{A_1, \ldots, A_k\}) \setminus \{A_{k+1} \}} P(C' \land \neg A_{k+1} \land D)$$

$$= \sum_{C' \in \text{Conj}(X) \setminus \{A_{k+1} \}} P(C' \land D).$$

Putting everything into (1), we have

$$P^*(\bigwedge_{i=1}^k L_i \land A_{k+1} \land D)$$

$$= P(\bigwedge_{i=1}^k L_i \land A_{k+1} \land D) + \sum_{C' \in \text{Conj}(X) \setminus \{A_{k+1} \}} P(C' \land D).$$

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 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.

### 4.2 Naive Update

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 $U_{na} : \mathcal{P} \times \{A, \neg A \mid A \in \text{Args}\} \rightarrow \mathcal{P}$ is defined by $U_{na}(P, L) = H_1(P, L)$. 
\(U_{na}\) satisfies all basic properties and STRICT-Satisfaction (and therefore also \(\epsilon\)-WEAK satisfaction).

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

*Proof.* Uniqueness and Completeness follow immediately from the fact that \(U_{na}\) yields one probability distribution by definition.

For idempotence, note that by definition of \(U_{na}\), we have for \(P^* = U_{na}(P, L)\) that 
\(P^*(X) = 0\) whenever \(X \models \neg L\). Therefore when applying \(U_{na}\) a second time to \(P\) with \(L\), no probability mass is shifted and 
\(U_{na}(P^*, L) = P^*\).

Order Invariance follows from commutativity of refinement functions, see [10], Proposition 8.

STRICT-Satisfaction follows immediately from observing that the naive update shifts all probability mass to the worlds that satisfy the literal \(L\) that we update with.

\[\Box\]

Naive update functions do not take the semantics of the argument graph into account and therefore can generally violate all rationality postulates that we introduced.

**Example 2.** Consider an argument graph over \(A, B\) with Attacks = \(\{A, B\}\). Let \(P\) be defined by \(P(\{B\}) = 1\) \(P\) satisfies RAT and COH since \(P(A) = 0\). However, for \(P' = U_{na}(P, A)\), we have \(P'((A, B)) = 1\). Hence, \(P'(A) = P'(B) = 1\) and RAT and COH are violated.

Next, consider \(P\) defined by \(P(\{A\}) = 1\). Then SFOU, FOU and OPT are satisfied. However, for \(P' = U_{na}(P, \neg A)\), we have \(P(\emptyset) = 1\). Hence, \(P'(A) = 0\) and SFOU, FOU and OPT are violated.

Finally, consider \(P\) defined by \(P(\{A\}) = 0.5\), \(P(\{B\}) = 0.5\). Then SOPT is satisfied. However, for \(P' = U_{na}(P, \neg B)\), we will have \(P'((A)) = 0.5\) and \(P'(\emptyset) = 0.5\). Therefore \(P'(B) = 0 < 1 - 0.5 = 1 - P(A)\) and SOPT is violated.

However, given an update literal over the argument \(A\), the naive update is guaranteed to be minimal with respect to \(d^{(A)}_{Jo}\) - in fact, the change with respect to \(d^{(A)}_{Jo}\) is 0.

**Proposition 2.** Let \(P \in \mathcal{P}\) and for some \(A \in \text{Args}\), let \(L \in \{A, \neg A\}\) be a literal. Then 
\(d^{(A)}_{Jo}(P, U_{na}(P, L)) = 0\).

*Proof.* We prove only the case \(L = A\) since the case for \(L = \neg A\) is analogous. Consider an arbitrary \(C \in \text{Conj}(\text{Args} \setminus \{A\})\) and let \(P^* = U_{na}(P, L)\). Since \(A \land C\) and \(\neg A \land C\) are complete conjunctions over \(\text{Args}\), we have by definition of \(U_{na}(P, L)\)

\[P^*(C) = P^*(A \land C) + P^*(\neg A \land C)\]

\[= (P(A \land C) + 1 \cdot P(\neg A \land C)) + (1 - 1) \cdot P(\neg A \land C)\]

\[= P(A \land C) + P(\neg A \land C) = P(C)\]

Hence, 
\(d^{(A)}_{Jo}(P, P^*) = \sum_{C \in \text{Conj}(\text{Args} \setminus \{A\})} |P(C) - P^*(C)| = 0\). \(\Box\)

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.
4.3 Trusting Update

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** $U_{tr} : \mathcal{P} \times \text{Args} \rightarrow \mathcal{P}$ is defined by $U_{tr}(P, A) = H_1(P, \Phi)$, where $\Phi = \{A\} \cup \{-C \mid (A, C) \in \text{Attacks}(G) \text{ or } (C, A) \in \text{Attacks}(G)\}$.

The trusting update has the same basic properties as the naive update, but additionally maintains some semantical constraints over the argument graph.

**Proposition 3.** $U_{tr}$ satisfies Uniqueness, Completeness, Idempotence, Order Invariance and STRICT-Satisfaction.

**Proof.** Uniqueness, Completeness follow immediately analogously to the naive update.

Idempotence follows also similarly with Lemma 2 that says that all probability mass is shifted to the conjunction of the update literal and the negated arguments that attack or are attacked by the update literal. Therefore, a second application can again cause no change in the probability distribution.

Order Invariance follows again from commutativity of refinement functions [10].

STRICT-Satisfaction follows again from observing that all probability mass is shifted to the worlds that satisfy the update argument. □

Before looking at R-consistency, let us note that the trusting update is again minimal with respect to the joint distance, this time defined with respect to the update argument and all of its attackers and attackees. Again, the joint distance is 0.

**Proposition 4.** Let $P \in \mathcal{P}$ and let $A \in \text{Args}$. Let $\Phi = \{A\} \cup \{-C \mid (A, C) \in \text{Attacks}(G) \text{ or } (C, A) \in \text{Attacks}(G)\}$. Then $d^\Phi_{jo}(P, U_{tr}(P, \Phi)) = 0$.

**Proof.** Consider an arbitrary $D \in \text{Conj}(\text{Args} \setminus \Phi)$ and let $P^* = U_{tr}(P, L)$. Let $C' = A \land \land_{L \in \Phi \setminus \{A\}} \neg L$. We have

$$P^*(D) = \sum_{C \in \text{Conj}(\Phi)} P^*(C \land D) = P^*(C' \land D) \sum_{C \in \text{Conj}(\Phi) \setminus \{C'\}} P^*(C \land D).$$

We know from Lemma 2 that

$$P^*(C' \land D) = P(C' \land D) + \sum_{C \in \text{Conj}(\Phi) \setminus \{C'\}} P(C \land D) = P(D).$$

and that $P^*(C \land D) = 0$ for all $C \in \text{Conj}(\Phi) \setminus \{C'\}$. Hence, $P^*(D) = P(D)$. Since $D \in \text{Conj}(\text{Args} \setminus \Phi)$ was arbitrary, we have $d^\Phi_{jo}(P, U_{tr}(P, \Phi)) = \sum_{D \in \text{Conj}(\text{Args} \setminus \Phi)} |P(D) - P^*(D)| = 0$ □

$U_{tr}$ maintains rationality and coherence.

**Proposition 5.** $U_{tr}$ satisfies R-Satisfaction for all $R \subseteq \{\text{RAT}, \text{COH}\}$. 
Proof. We have to show that if \( P \models R \) before the update, \( P' \models R \) for \( P' = U_{\text{tr}}(P, A) \). Proposition 4 implies that \( P(B) \neq P'(B) \) for some \( B \in \text{Args} \setminus \{A\} \) is only possible if \((A, B)\) or \((B, A)\) are in \text{Attacks}. In both cases, we will have \( P'(A) = 1 \) and \( P'(B) = 0 \). By looking at the definitions of RAT and COH, we can see that they cannot be violated after the update.

However, the remaining semantical constraints can be violated after updates with \( U_{\text{tr}} \).

Example 3. Consider an argument graph over \( A, B \) with \text{Attacks} = \{(A, B)\}. Then for each \( P \in \mathcal{P} \), \( P' = U_{\text{tr}}(P, B) \) will satisfy \( P'(B) = 1 \) and \( P'(A) = 0 \) (because \( A \) attacks \( B \)). Since \( A^{-} = \emptyset \), \( P' \) will violate SFOU, FOU and OPT.

For R-SOPT, consider an argument graph over \( A, B, C \) with \text{Attacks} = \{(A, B), (B, C)\}. Let \( P \) be defined by \( P(\{B\}) = 1 \). Then \( P \) satisfies SOPT. \( P' = U_{\text{tr}}(P, A) \) can be obtained by first computing \( H_1(P, A) = P_1 \), where \( P_1(\{A, B\}) = 1 \) and then computing \( H_1(P_1, \neg B) = P' \). Note that \( P'(\{A\}) = 1 \). In particular, \( P'(B) = P'(C) = 0 \) and \( P'(C) < 1 - P'(B) \), i.e., SOPT is violated.

4.4 Strict Update

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 \( U_{\text{st}} : \mathcal{P} \times \text{Args} \rightarrow \mathcal{P} \).

For \( A \in \text{Args} \), let \( \Phi = \{A\} \cup \{\neg C \mid (A, C) \in \text{Attacks} \} \) and let the constraint \( C(P) \) be true iff for all \((B, A) \in \text{Attacks}, P(B) \leq 0.5 \). Then \( U_{\text{st}}(P, A) = P^* \) where

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

The strict update function satisfies all basic properties except Order Invariance.

Proposition 6. \( U_{\text{st}} \) satisfies Uniqueness, Completeness and Idempotence.

Proof. Uniqueness, Completeness follow immediately analogously to the naive update.

For Idempotence, we have to look at two cases. If there is a \((B, A) \in \text{Attacks} \) such that \( P(B) > 0.5 \), \( P \) will remain unchanged, i.e., \( P = U_{\text{st}}(P, A) \). In particular \( U_{\text{st}}(U_{\text{st}}(P, A), A) = U_{\text{st}}(P, A) \). If \( P(B) \leq 0.5 \) for all \((B, A) \in \text{Attacks} \), Idempotence can be checked analogously to the trusting update.

The following example illustrates how Order Invariance can be violated by the strict update.

Example 4. Consider an argument graph over \( A, B \) with \text{Attacks} = \{(A, B), (B, A)\}.

Let \( P \) be defined by \( P(\emptyset) = 1, P_A = U_{\text{st}}(P, A) \) is completely described by \( P_A(\{A\}) = 1 \) and \( P_B = U_{\text{st}}(P, B) \) by \( P_B(\{B\}) = 1 \). In particular, \( U_{\text{st}}(P_A, B) = P_A \neq P_B = U_{\text{st}}(P_B, A) \). Hence, Order Invariance is violated.
Due to its conditional update, the strict update can violate \( \epsilon \)-WEAK-consistency (and therefore also STRICT-consistency).

**Example 5.** Consider an argument graph over \( A, B \) with \( \text{Attacks} = \{(A, B)\} \). Let \( P \)
be defined by \( P(\{A\}) = 1 \). Then \( P = U_{\text{st}}(P, B) \) and \( P(B) = 0 < 0.5 \). Hence, \( \epsilon \)-WEAK-consistency is violated for all \( \epsilon \in (0, 0.5) \).

Of course, the violation is intended here. We could introduce a conditional satisfaction condition that is satisfied by \( U_{\text{st}} \), but let us not do so here in order to keep things simple.

\( U_{\text{st}} \) again guarantees joint distance 0, this time with respect to the update argument and all of its attackers.

**Proposition 7.** Let \( P \in \mathcal{P} \) and let \( A \in \text{Args} \). Let \( S = \{A\} \cup \{\neg C \mid (A, C) \in \text{Attacks}(C)\} \). Then \( d_{\text{Jo}}^A(P, U_{\text{st}}(P, S)) = 0 \).

**Proof.** There are two cases. If there is a \((B, A) \in \text{Attacks} \) such that \( P(B) > 0.5 \), \( P \) will remain unchanged and therefore \( d_{\text{Jo}}^A(P, U_{\text{st}}(P, S)) = 0 \). Otherwise, the proof is completely analogous to the proof of Proposition 4. \( \Box \)

\( U_{\text{st}} \) satisfies all semantical constraints but \( \text{OPT} \) and \( \text{SOPT} \).

**Proposition 8.** \( U_{\text{st}} \) satisfies R-Satisfaction for all \( R \subseteq \{\text{RAT, COH, SFOU, FOU}\} \).

**Proof.** We have to show that if \( P \models R \) before the update, \( P' \models R \) for \( P' = U_{\text{st}}(P, A) \). If there is a \((B, A) \in \text{Attacks} \) such that \( P(B) > 0.5 \), \( P \) will remain unchanged and the claim is true. Otherwise, Proposition 7 implies that \( P(B) \neq P'(B) \) for some \( B \in \text{Args} \setminus \{A\} \) is only possible if \((A, B) \in \text{Attacks} \). We will then have \( P'(A) = 1 \) and \( P'(B) = 0 \). By looking at the definitions of \( \text{RAT, COH} \), we can see that they cannot be violated after the update. Since only the probabilities of \( A \) and attacked arguments \( B \) are changed \( \text{SFOU} \) and \( \text{FOU} \) cannot be violated after the update either. \( \Box \)

The following shows how \( \text{R-OPT} \) and \( \text{R-SOPT} \) can be violated by \( U_{\text{st}} \).

**Example 6.** Consider an argument graph over \( A, B, C \) with \( \text{Attacks} = \{(A, B), (B, C), (C, A)\} \). Let \( P \) be defined by \( P(\{A, B, C\}) = 0.4, P(\{A, B\}) = 0.3, P(\emptyset) = 0.3 \).

Then \( P(A) = P(B) = 0.7, P(C) = 0.4 \) and \( \text{OPT} \) and \( \text{SOPT} \) are satisfied. Since \( P(C) \leq 0.5 \), we can update \( A \). \( P' = U_{\text{st}}(P, A) \) can be obtained by computing \( H_1(P, A) = P_1 \), where \( P_1(\{A, B, C\}) = 0.4, P_1(\{A, B\}) = 0.3, P_1(\{A\}) = 0.3 \), and then computing \( H_1(P_1, \neg B) = P' \), where \( P'(\{A, C\}) = 0.4, P'(\{A\}) = 0.6 \). In particular, we have \( P'(B) = 0 \) and \( P'(C) = 0.4 \) and therefore \( P'(C) < 1 - P'(B) \). Hence, \( \text{OPT} \) and \( \text{SOPT} \) are violated after the update.

## 5 R-S-d Update Functions

### 5.1 Definition and Analysis

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 \( R \subseteq \{ \text{RAT, COH, SFOUT, FOU, SOPT, OPT} \} \), \( S \in \{ \text{STRICT, } \epsilon\text{-WEAK} \} \) and \( d \in \{ d_1, d_2, d_{\infty}, d^X_{\alpha}, d^\Phi_{\alpha} \} \). An \( R\text{-S-d Update Function} \) \( U_{R,S,d} : \mathcal{P} \times \text{Form} \rightarrow 2^\mathcal{P} \) is defined by
\[
U_{R,S,d}(P, F) = \arg \min_{P' \in \text{Mod}_{R,S,d}(F)} d(P, P').
\]

R-S-d Update functions can easily be parameterized to satisfy our semantical properties and minimal change properties.

Proposition 9. For all \( R \subseteq \{ \text{RAT, COH, SFOUT, FOU, SOPT, OPT} \} \), \( S \in \{ \text{STRICT, } \epsilon\text{-WEAK} \} \) and \( d \in \{ d_1, d_2, d_{\infty}, d_m, d_{KL}, d^X_{\alpha}, d^\Phi_{\alpha} \} \), \( U_{R,S,d} \) satisfies R-consistency, S-consistency and R-S-d-minimality.

Actually, \( U_{R,S,d} \) satisfies a stronger property than R-consistency. Even if \( P \not\models R \), we will have \( U_{R,S,d}(P, F) \models R \). This follows again immediately from the definition. This feature of R-S-d update functions allows us to initialize epistemic states that satisfy certain semantical properties. For instance, given a formula \( F \), we could consider the uniform distribution \( P_0 \in \mathcal{P} \) and then compute \( U_{R,S,d}(P_0, F) \). However, initializing the epistemic states in this way only works if \( |U_{R,S,d}(P_0, F)| = 1 \). Indeed, R-S-d update functions are not unconditionally complete because we might have inconsistencies. Let us first note some analytic properties of \( U_{R,S,d}(P, F) \).

Lemma 3. For each \( R \subseteq \{ \text{COH, SFOUT, FOU, SOPT, OPT} \} \) (we left out RAT), \( S \in \{ \text{STRICT, } \epsilon\text{-WEAK} \} \) and \( d \in \{ d_1, d_2, d_{\infty}, d_m, d_{KL}, d^X_{\alpha}, d^\Phi_{\alpha} \} \), 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 \( \text{Mod}_{R,S}(F) \) is non-empty.

If \( R \) includes RAT, \( U_{R,S,d}(P, F) \) will be non-empty and compact whenever \( \text{Mod}_{R,S}(F) \) is non-empty.

Proof. The constraints corresponding to COH, SFOUT, FOU, SOPT, OPT, JUS and STRICT, \( \epsilon\text{-WEAK} \) are linear inequality constraints over probability distributions. If we fix one argument, all distance measures are convex and continuous in the other argument. Therefore, computing \( U_{R,S,d}(P, F) \) corresponds to a convex combination problem and a minimum exists whenever \( \text{Mod}_{R,S}(F) \neq \emptyset \). Linearity of the constraints implies that \( \text{Mod}_{R,S}(F) \) is convex. Therefore, each convex combination of probability distributions \( P_1, P_2 \) that minimize \( d(P, \cdot) \) over \( \text{Mod}_{R,S}(F) \), will also be in \( U_{R,S,d}(P, F) \). In particular, each convex combination \( \lambda P_1 + (1 - \lambda) P_2 \) will be again minimize \( d(P, \cdot) \) because \( d(P, \lambda P_1 + (1 - \lambda) P_2) = \lambda d(P, P_1) + (1 - \lambda) d(P, P_2) \). Hence, \( U_{R,S,d}(P, F) \) is convex. Linearity of the constraints also implies that \( \text{Mod}_{R,S}(F) \) is closed. Therefore, the limit of each sequence of probability distributions that minimize \( d(P, \cdot) \) will be in \( \text{Mod}_{R,S}(F) \). By continuity of \( d \), the limit will also minimize \( d(P, \cdot) \). Therefore, \( U_{R,S,d}(P, F) \) is closed. Since \( \mathcal{P} \) is bounded, \( U_{R,S,d}(P, F) \) is in particular compact.

As explained in [11], RAT corresponds to the union of two linear constraints \( P(A) \leq 0.5 \) and \( P(B) \leq 0.5 \) (this can be seen by rewriting the implication in RAT as a disjunction). Since closed sets are closed under union, \( \text{Mod}_{R,S}(F) \) will still be closed when \( R \) includes RAT. Therefore, continuity of \( d \) implies that a minimum exists whenever \( \text{Mod}_{R,S}(F) \) is non-empty. Compactness of \( U_{R,S,d}(P, F) \) follows as before. \( \square \)
From Lemma 3, we get the following completeness guarantee

**Corollary 1.** For all \( R \subseteq \{\text{COH}, \text{SFOU}, \text{FOU}, \text{SOPT}, \text{OPT}, \text{RAT}\} \), \( S \in \{\text{STRICT}, \epsilon\text{-WEAK}\} \) and \( d \in \{d_1, d_2, d_\infty, d_{KL}, d_{AM}, d_{\Phi_1}\} \), \( U_{R,S,d}(P,F) \geq 1 \) whenever \( F \) is \( R\text{-S-consistent} \).

**Proof.** If \( F \) is \( R\text{-S-consistent} \), \( \text{Mod}_{R,S}(F) \neq \emptyset \) and the claim follows from Lemma 3.

\( U_{R,S,d} \) will not necessarily yield a unique probability distribution. However, we have Uniqueness for the Least Squares distance and the KL-divergence.

**Proposition 10.** For each \( R \subseteq \{\text{COH}, \text{SFOU}, \text{FOU}, \text{SOPT}, \text{OPT}\} \) (we left out \( \text{RAT} \)) and \( S \in \{\text{STRICT}, \epsilon\text{-WEAK}\} \) and \( d \in \{d_2, d_{KL}\} \), \( |U_{R,S,d}(P,F)| \leq 1 \).

**Proof.** Both the Least Squares distance and the KL divergence are strictly convex (this can be checked by noting that the Hessian matrix of both functions is a diagonal matrix with positive diagonal entries and therefore positive definite). As explained before, \( \text{Mod}_{R,S}(F) \) is convex. The claim follows from the fact that the minimum of a strictly convex function on a convex set is unique.

**Remark 2.** Strictly speaking, we need to impose a stronger condition on the KL-divergence, namely that \( \text{Mod}_{R,S}(F) \) contains an element \( P' \) that is ‘absolutely continuous’ with respect to the prior distribution \( P \), that is, \( P'(X) = 0 \) whenever \( P(X) = 0 \). Roughly speaking, the KL-divergence is \( \infty \) for all elements in \( \text{Mod}_{R,S}(F) \) if there is no such \( P' \). In order to keep things simple, we do not discuss the absolute continuity assumption and its implications here, see [14] for a thorough discussion.

For the other change measures, we can have an infinite number of minimal solutions. We could select a best one with respect to some rationality criterion like the principle of maximum entropy to guarantee Uniqueness in these cases. However, we are not going to do so here.

Tautology can only be satisfied if we have \( P \in \text{Mod}_{R,S}(\top) \) for the prior distribution \( P \) and in particular only if the constraints in \( R \) and \( S \) are compatible. Furthermore, it can be violated when using the atomic and joint distance because with respect to these, there can be an infinite number of probability distributions that have distance 0 to \( P \).

**Proposition 11.** For all \( R \subseteq \{\text{COH}, \text{SFOU}, \text{FOU}, \text{SOPT}, \text{OPT}, \text{RAT}\} \), \( S \in \{\text{STRICT}, \epsilon\text{-WEAK}\} \) and \( d \in \{d_1, d_2, d_\infty, d_{KL}\} \) (we left out \( d_{AM}, d_{\Phi_1}\)), if \( P \in \text{Mod}_{R,S}(\top) \neq \emptyset \), then \( U_{R,S,d}(P,\top) = \{P\} \).

**Proof.** Since \( P \in \text{Mod}_{R,S}(\top) \neq \emptyset \), \( P \) is an optimal solution with distance 0. \( P \) is the only solution with distance 0 because \( d_1, d_2, d_\infty \) and \( d_{KL} \) are definite.

Contradiction and Representation Invariance are satisfied in most cases.

**Proposition 12.** For all \( R \subseteq \{\text{COH}, \text{SFOU}, \text{FOU}, \text{SOPT}, \text{OPT}, \text{RAT}\} \), \( S \in \{\text{STRICT}, \epsilon\text{-WEAK}\} \) and \( d \in \{d_1, d_2, d_\infty, d_{KL}\} \), \( U_{R,S,d} \) satisfies Contradiction and Representation Invariance.
Table 3. Illustration of R-S-d updates with \( R_1 = \{COH\}, R_2 = \{COH, SOPT\}, S = STRICT \) and \( d = d_2 \).

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>( P )</th>
<th>( U_{R_1, S, d}(P, A) )</th>
<th>( U_{R_1, S, d}(P, B) )</th>
<th>( U_{R_2, S, d}(P, A) )</th>
<th>( U_{R_2, S, d}(P, B) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.17</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.49</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.1</td>
<td>0.55</td>
<td>0</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.12</td>
</tr>
</tbody>
</table>

**Proof.** Contradiction follows from the fact that \( P(\bot) = 0 \) for all \( P \in \mathcal{P} \). Hence, \( Mod_{R,S}(\bot) = \emptyset \) and therefore \( U_{R,S,d}(P, \bot) = \emptyset \).

Representation invariance follows from the fact that \( F \equiv G \) implies \( P(F) = P(G) \). Therefore, \( Mod_{R,S}(F) = Mod_{R,S}(G) \) and \( U_{R,S,d}(P, F) = U_{R,S,d}(P, G) \). \( \square \)

Idempotence can be violated again when using the atomic and joint distance because of their violation of definiteness. However, we have the following result.

**Proposition 13.** For all \( R \subseteq \{COH, SFOU, FOU, SOPT, OPT, RAT\} \), \( S \in \{STRICT, \epsilon\text{-WEAK}\} \) and \( d \in \{d_1, d_2, d_\infty, d_{KL}\} \) (we left out \( d_\Phi, d_{KL}^\Phi \)), \( U_{R,S,d} \) satisfies Idempotence.

**Proof.** If \( U_{R,S,d}(P, F) = \{P^*\} \), then \( P^* \in Mod_{R,S}(F) \). Hence, \( U_{R,S,d}(P^*, F) = \{P^*\} \) because by definiteness of \( d_1, d_2, d_\infty \) and \( d_{KL} \), \( P^* \) is the only element in \( \mathcal{P} \) with distance 0 to \( P^* \). \( \square \)

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 7.** 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 \( P_2(\{A\}) = 0.25, P_2(\{A, B\}) = 0.75 \).

### 5.2 Relationships between Refinement-Based Update Functions and R-S-d Update Functions

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 14.** 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 \( A \in \text{Args} \). If there is a \( P^* \in Mod_{R,STRICT}(L) \) such that \( d_{j_0}^{\{A\}}(P, P^*) = 0 \) then \( U_{R,STRICT,d_{j_0}^{\{A\}}}(P, L) = \{U_{\text{na}}(P, L)\} \).
Proof. Without loss of generality, assume \( L = A \) (for \( L = \neg A \), the proof is analogous). Since \( d^A_{j_0}(P, P^*) = 0 \), \( P^* \in U_{R, STRICT, d^A_{j_0}}(P, L) \) and \( d^A_{j_0}(P, P') = 0 \) for all \( P' \in U_{R, STRICT, d^A_{j_0}}(P, L) \). In particular, for all complete conjunctions \( C \in \text{Conj}(\text{Args} \setminus \{ A \}) \), we must have \( P'(C) = P(C) \) for otherwise the distance could not be 0. Furthermore, \( P'(\neg A_L \land C) = 0 \) because \( P' \) strictly satisfies \( L \) (\( P'(A_L) = 1 \)) and \( P'(\neg A_L \land C) \leq P'(\neg A_L) = 1 - P'(A_L) = 0 \). Therefore,

\[
P'(A_L \land C) = P'(C) - P'(\neg A_L \land C) = P(C) - 0 = P(A_L \land C) + P(\neg A_L \land C).
\]

Each \( A_L \land C = A_L \land \bigwedge_{A \in \text{Args} \setminus A_L} A^b_a \) is a complete conjunction over \( \text{Args} \) and corresponds to exactly one interpretation \( X \subseteq 2^{\text{Args}} \) as explained before. Therefore, \( P' \) equals \( U_{\text{na}}(P, L) \) for all interpretations that contain \( A_L \). For all interpretations that do not contain \( A_L \), we have \( P'(\neg A_L \land C) = 0 \) as discussed before. Hence, \( P' \) also equals \( U_{\text{na}}(P, L) \) for all interpretations that do not contain \( A_L \). Hence, \( P' = U_{\text{na}}(P, L) \). \( \Box \)

Remark 3. Note that if there is no \( P^* \in \text{Mod}_{R, STRICT}(L) \) such that \( d^A_{j_0}(P, P^*) = 0 \), 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, \( U_{R, STRICT, d^A_{j_0}} \) agrees with \( U_{\text{na}} \) whenever \( U_{\text{na}} \) is consistent with \( R \). Otherwise, \( U_{R, STRICT, d^A_{j_0}} \) 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 2. \( U_{\emptyset, STRICT, d^A_{j_0}}(P, L) = \{ U_{\text{na}}(P, L) \} \).

Proof. We have \( U_{\text{na}}(P, L) \in \text{Mod}_{\emptyset, STRICT}(L) \) because for \( P' = U_{\text{na}}(P, L) \), we have \( P'(L) = 1 \). We know that \( d^A_{j_0}(P, U_{\text{na}}(P, L)) = 0 \) from Proposition 2. Hence, the claim follows from Proposition 14.

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

Proposition 15. Consider an arbitrary set of semantical constraints \( R \subseteq \{ \text{RAT}, \text{COH}, \text{SFOU}, \text{FOU}, \text{SOPT}, \text{OPT} \} \), a probability distribution \( P \in \mathcal{P} \) and let \( L \in \{ A, \neg A \} \) be a literal for some \( A \in \text{Args} \). Let \( X' = \{ C \mid (A, C) \in \text{Attacks(G)} \) or \( (C, A) \in \text{Attacks(G)} \} \) and \( X = \{ A \} \cup X' \). If there is a \( P^* \in \text{Mod}_{R, STRICT}(L) \) such that \( d_{j_0}^X(P, P^*) = 0 \) then \( U_{R, STRICT, d_{j_0}^X}(P, L \land \bigwedge_{C \in X'} \neg C) = \{ U_{\text{tr}}(P, L) \} \).

Proof. Without loss of generality, assume \( L = A \) (for \( L = \neg A \), the proof is analogous). Since \( d_{j_0}^X(P, P^*) = 0 \), \( P^* \in U_{R, STRICT, d_{j_0}^X}(P, L \land \bigwedge_{C \in X'} \neg C) \) and \( d_{j_0}^X(P, P') = 0 \) for all \( P' \in U_{R, STRICT, d_{j_0}^X}(P, L \land \bigwedge_{C \in X'} \neg C) \). In particular, for all complete conjunctions \( D \in \text{Conj}(\text{Args} \setminus X) \), we must have \( P'(D) = P(D) \) for otherwise the distance could not be 0. Furthermore, for all complete conjunctions \( C \in \text{Conj}(X) \setminus \{ L \land \bigwedge_{C \in X'} \neg C \} \), \( P'(C \land D) = 0 \) because \( P' \) strictly satisfies \( L \land \bigwedge_{C \in X'} \neg C \).
\( P'(L \land \bigwedge_{C \in X'} \neg C) = 1 \) and \( P'(C \land D) \leq P'(C) \leq 1 - P'(L \land \bigwedge_{C \in X'} \neg C) = 0 \). Therefore,

\[
P'(L \land \bigwedge_{C \in X'} \neg C \land D) = P'(D) - \sum_{C \in \text{Conj}(X) \setminus \{L \land \bigwedge_{C \in X'} \neg C\}} P'(C \land D)
\]

\[
= P(D) - 0 = P(L \land \bigwedge_{C \in X'} \neg C \land D) + \sum_{C \in \text{Conj}(X) \setminus \{L \land \bigwedge_{C \in X'} \neg C\}} P(C \land D).
\]

From Lemma 2 we can see that \( P' = \{U_{\text{tr}}(P, L)\} \).

**Corollary 3.** \( U_{\emptyset, \text{STRICT}, d_{J_{na}}}(P, L \land \bigwedge_{C \in X'} \neg C) = \{U_{\text{tr}}(P, L)\} \).

**Proof.** Let \( P' = U_{\text{tr}}(P, L) \). Lemma 2 implies that \( P'(L \land \bigwedge_{C \in X'} \neg C) = 1 \) because all probability mass is shifted to the models of \( L \land \bigwedge_{C \in X'} \neg C \). Hence, \( U_{\text{tr}}(P, L) \in \text{Mod}_{\emptyset, \text{STRICT}}(L) \). We know that \( d_{J_{na}}^2(P, U_{\text{na}}(P, L)) = 0 \) from Proposition 4. Hence, the claim follows from Proposition 15.

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 differentiation that is used for the strict update.

6 Conclusions and Future Work

Most proposals for dialogical argumentation focus on protocols (e.g., [16], [17], [5], [2]) with strategies being under-developed. See [20] 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. [18, 6, 7, 19]). 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 persuadee 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 [13] or select solutions that violate the constraints in a minimal way [3, 15]. 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